Investigation of *Mal d 1* Allelic Variants and Phylogenetic Diversity in Contemporary and Historical Polish Apple Cultivars

Kamila Lucia Bokszczanin¹, Andrzej Antoni Przybyla², Nicolas Krezdorn¹, Peter Winter¹ & Marcin Filipecki³

¹ GenXPro GmbH, Frankfurt am Main, Germany
² Department of Pomology, Faculty of Horticulture, Biotechnology and Landscape Architecture, Warsaw University of Life Sciences, Warsaw, Poland
³ Department of Plant Genetics, Breeding and Biotechnology, Faculty of Horticulture, Biotechnology and Landscape Architecture, Warsaw University of Life Sciences, Warsaw, Poland

Correspondence: Kamila Lucia Bokszczanin, GenXPro GmbH, Frankfurt am Main, Germany. E-mail: kbokszczanin@genxpro.de

Received: July 26, 2015  Accepted: August 25, 2015  Online Published: October 15, 2015
doi:10.5539/jas.v7n11p115  URL: http://dx.doi.org/10.5539/jas.v7n11p115

Abstract

Allergy is considered as the most common health problem of XXI century. Apple allergenicity is of great importance and at present the only therapy for food allergy is avoidance. Fruits are important components of a healthy diet, and therefore complete avoidance of apple and related fruits like pear, peach, and cherry can have a significant negative impact since deprives the patient’s diet from important sources of vitamins, minerals and fibers. In order to provide a contribution towards a better understanding of the genetics of the family of *Mal d 1* gene, the major apple allergen, and to establish the basis for further research on allelic diversity among cultivars in relation to variation in allergenicity we evaluated *Mal d 1* gene variability in contemporary and historical Polish apple cultivars (*Malus × domestica* Borkh.). We identified new *Mal d 1* alleles and further assessed phylogenetic distance between *Mal d 1* variants and antigens derived from other food sources. We also predicted possible cross-reactions between allergens originated from *Malus × domestica* and other plant taxa as well as showed potential binding sites of apple allergens to the antibody. Thanks to multiple sequence alignment of the amino acid sequences limited to residues which serve as interaction partners for the antibody we were able to more precisely establish phylogenetic distance between analyzed antigens in regards to the antigen epitope residues.

Keywords: allergenicity, *Malus × domestica*, apple allergens genetic variability

1. Introduction

Apple fruits are considered to be one of the major sources of phenol compounds due to their widespread consumption in many countries worldwide and availability on the market throughout the year. An apple fruit dietary properties are related to the fermentable fiber content which decreases transit time, increases bulk, lowers pH, and produces potentially anticarcinogenic short-chain fatty acids (Word Cancer Research Fund/American Institute for Cancer Research). Thus it has been suggested that greater intake of apple could contribute to improved health. However apples are also a considerable source of allergens and among the plant food fruits belonging to the *Rosaceae* family are the most frequently responsible for allergic reactions. It has been documented, that approx. 2% of North and Central European population is allergic to apples and nowadays the only therapy for food allergy is avoidance (Kootstra et al., 2007).

Allergy has been described as the epidemic of the 21st century, affecting up to 40% of the general population of the developed countries (The International Study of Asthma and Allergies in Childhood (ISAAC). Clinical studies performed in Europe revealed that 2-4% of the adult and 6-8% of the children population is affected by food allergies what is comparable to 3.5-4% of the US population suffering from IgE-mediated food allergy (Gupta et al., 2011). It is well established that up to 70% of birch pollen allergic patients in Northern and Central Europe display predominantly mild allergic symptoms when eating plant food such as apple, peach, nuts, celery or spices. In Central and Northern Europe *Rosaceae* food allergy, mainly apple allergy, is associated with birch pollinosis and mild localized oropharyngeal symptoms (oral allergy syndrome, OAS) with symptoms appearing
generally within 5-15 min. after consuming raw apple with itching, tingling or swelling of lips, mouth, tongue or throat. Allergy to apple without sensitization to birch pollen is extremely rare in these areas. In contrast, in Southern Europe, where birch trees are virtually absent, allergies to both apple and related *Rosaceae* fruits (like peach and plum) which affect pollen and non-pollen allergic patients (Fernandez-Rivas et al., 1997), are manifested by mild OAS and frequently by severe systemic reactions.

Generally pollen food syndrome i.e. OAS results from cross-reactivity. The phenomenon of allergen cross-reactivity occurs when IgE antibodies rose against one allergen (e.g. from birch pollen) binds or recognizes a similar protein from another source. This homology of proteins derived from close-related and unrelated plant species is a result of similarity in the 3D folding structural features. The current model for cross-reactivity is that the major allergens with ≥ 75% sequence identity have common molecular structures (i.e. epitopes) on their surfaces, which are recognized by the same patient IgE antibodies (Ipsen et al., 1992). Also, more distantly related species, such as apple, contain molecules homologous to Bet v 1 (Ebner et al., 1991; Aalberse et al., 2001). Bet v 1 from birch and other allergens from species within *Fagales* order (alder, hornbeam, chestnut, hazelnut, white oak) as well as fruits belonging to the *Rosaceae* family are PR-10 homologs (pathogenesis-related class 10 proteins) which increase in plants, when they are exposed to stress. IgE antibodies specific for Bet v 1 have been shown to cross-react with homologous proteins identified in different fruits and vegetables such as apple, pear, sweet cherry, apricot, hazelnut, strawberry, peanut, celery, carrot, soybean, potato, and mungbean.

It is considered that differences in apple allergenicity are determined by patient-dependent sensitivitiy, apple cv, degree of fruit maturity, and fruit storage conditions (Vieths et al., 1994; Asero et al., 2006; Botton et al., 2008). There is intrinsic importance of binding capacities of Bet v 1 specific IgE antibodies to Mal d 1 isoforms. Also genetic variation of *Mal d 1* gene and its expression pattern in the different cvs determines apple cv allergenicity.

In apple four allergens have been identified so far: *Mal d 1*, *Mal d 2*, *Mal d 3*, and *Mal d 4*. *Mal d 1* which is considered as a major apple allergen causing allergic symptoms in birch-pollen sensitized patients is localized in both pulp and peel of the apple. More severe reactions are related to *Mal d 3* present in the peel. *Mal d 1*, the major allergen in apple shares 55–65% amino acid sequence identity with Bet v 1 (Holm et al., 2011), and both allergens possess conformational epitopes composed of discontinuous sections of antigen amino acid sequence recognized by the immune system.

At present the only therapy for food allergy is avoidance. Fruits are important components of a healthy diet, and therefore complete avoidance of apple and related fruits like pear, peach, cherry and nuts like hazelnut can have a significant negative impact since deprives the patient’s diet from important sources of vitamins, minerals, and fibers. Patients allergic to apples show great interest in consuming apple fruit, so it is justifiable to deepen research of new apple sources that can potentially be introduced on the market as hypoallergenic cultivars.

Here we aimed at evaluation of *Mal d 1* gene allelic variability in contemporary and historical Polish apple cultivars (*Malus × domestica* Borkh.). A broad variety of cultivars with a variable degree of allergenicity is grown in various European countries. For instance ‘Santańa’ and ‘Elise’ are considered as low allergenic apple cultivars as it has been proven in studies performed in the Netherlands, Great Britain, Switzerland, and northern Italy (Vlieg-Boerstra et al. 2013). Polish fruit production is dominated with apples which are number one on Polish market. Commercial apple production in the recent years amounted to 1.5–2.1 million t/year.

Via identification of already known as well as novel *Mal d 1* alleles present in Polish cultivars, investigation of the phylogenetic distances between all known *Mal d 1* variants and antigens derived from other food sources anticipated to cross-react as well as prediction of potential binding sites of *Mal d 1* variants with the antibody, our research contributes to better understanding the genetics of the *Mal d 1* gene family. Further our results extent the knowledge on *Malus* germplasm in relation to major apple allergen and provide relevant information for the breeders aiming at development of cultivars with the reduced allergenicity.

2. Materials and Methods

2.1 Cultivars Used for DNA Isolation and Genomic Cloning

*Apple* (*Malus × domestica* Borkh.) cultivars ‘Kantowka Gdanska’ (‘Danziger Kantapfel’), ‘Kosztela’, ‘Ligol’, ‘Redkroft’, and ‘Odra’ were used for genomic PCR cloning and sequencing. Leaves of aforementioned cultivars were harvested in the spring, lyophilized and stored at -80 °C until use.

‘Ligol’, ‘Redkroft’ and ‘Odra’ belong to Polish contemporary apple cultivars developed by Przybyla (Przybyla et al. 2009) at Research Institute of Horticulture in Skierniewice. ‘Kantowka Gdanska’ (first time described in 1760) has been developed in the vicinity of Gdansk in Poland while ‘Kosztela’ has been cultivated in Poland since
XVIth century.

2.2 DNA Isolation and Genomic PCR

Total genomic DNA was isolated from young lyophilized leaves using DNeasy Plant kit (Qiagen) according to manufacturer’s protocol. PCR analysis was carried out using hot-start proofreading enzyme with terminal transferase activity (extra A addition) followed by incubation with Taq Polymerase to provide assurance of efficient A-tailing facilitated subsequent cloning. Four primer pairs (Table 1) were used to obtain Mal d 1.01, Mal d 1.02, Mal d 1.04, and Mal d 1.06A-C gene amplicons. PCR was performed in a total volume of 20 µl containing 2.5 units/µl of HotStar HiFidelity DNA Polymerase (Qiagen), 1x HotStar HiFidelity PCR Buffer (with 0.2 mM dNTPs) (Qiagen), 1 µM of each primer, and 25–50 ng of genomic DNA. PCR amplification was performed according to the conditions described by Gao et al. (2005a) with initial activation step (5 min at 95 °C), and verified on 1% agarose gel. PCR fragments were A-tailed by adding 0.3 µl of Taq Polymerase (Qiagen) to each reaction mixture and subsequent incubating at 70 °C for 20 min. HiFidelity PCR buffer (enriched with SB Factor) ensures that the template DNA, which is often degraded by the 3’→5’ exonuclease activity of high-fidelity enzymes, remains intact. The preoptimized formulation provides reliable amplification of specific PCR products with a very low error rate by promoting a high ratio of specific-to-nonspecific primer binding during the annealing step in each PCR cycle. We used polymerase that prevents degradation of the template and to minimize enzyme-generated recombinant molecules produced in later cycles. It allowed avoiding either possible enzyme-directed degradation of the PCR product or formation of mismatched heteroduplexes, both of which occur after the maximal DNA concentration is reached (plateau phase), and increase with further cycling (over-amplification).

Table 1. Cloning primers and PCR conditions

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primersa</th>
<th>Proofreading polymeraseb Tm/cycles</th>
<th>Reference sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mal d 1.01</td>
<td>F:ATCTCCAACACAATACCTCAAC</td>
<td>58/25</td>
<td>AY789236</td>
</tr>
<tr>
<td></td>
<td>R:AAAGCCACACAACTTCCGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mal d 1.02</td>
<td>F:CATCCTTGGTAGTTTC</td>
<td>52/25</td>
<td>AY789239</td>
</tr>
<tr>
<td></td>
<td>R:ACCATAGAAACATATTAATTTAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mal d 1.04</td>
<td>F:CGTAGTTGGACAAGTGTTAGT</td>
<td>58/30</td>
<td>AY789242</td>
</tr>
<tr>
<td></td>
<td>R:AGGGTAACACACAAAAATTACATG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mal d 1.06A-C</td>
<td>F:CATGGGTGTCCTCACATCGGAC</td>
<td>55/25</td>
<td>AY789248</td>
</tr>
<tr>
<td></td>
<td>R:TTAGTTGGACCTACAGGATTG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. aPrimers for Mal d 1.04 adopted from Gao et al. (2008), others adopted from Gao et al. (2005 a, b, and c); bHotStar HiFidelity DNA Polymerase (Qiagen).

2.3 Genomic Cloning and Sequencing of Mal d 1.01, Mal d 1.02, Mal d 1.04 and Mal d 1.06 Genes

PCR products were ligated into the pCR®II-TOPO® vector (Invitrogen) and used to transform JM109 High Efficiency Competent Cells (Promega) according to the protocol recommended by the supplier. For each fragment, 8-16 white colonies were overnight subcultured, harvested with centrifugation and subjected to isolation with GenElute Plasmid Miniprep Kit (Sigma-Aldrich) according to manufacturer’s protocol. Recombinant plasmids were submitted for Sanger DNA sequencing with reverse and forward M13 primers. The nucleotide sequence of the cloned PCR products was determined by dideoxy chain termination using the BIG DYE Terminator Cycle Sequencing v1.1 DNA Sequencing Kit (Applied Biosystems). The reaction products were separated on 3730 DNA Analyzer (Applied Biosystems). All cloned PCR products were resequenced.

In order to validate allelic variants derived from Sanger-based sequencing (SBS), PCR amplified products obtained using primers encompassing variable regions of Mal d 1 alleles were subjected to 454 sequencing. Mapping using Novoalign was used to verify potential allelic variants.

2.4 Mapping

Mapping using Novoalign (version: V3.00.05) was used to verify potential allelic variants. Therefor the 454 sequences were mapped to the Sanger sequences. To identify the chromosomal location of the alleles we aligned
to the *M. x domestica* genome v1.0 (http://www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0).

2.5 Multiple Nucleotide Sequence Alignment

Allelic variants identified in analyzed cultivars were aligned using COBALT server, NCBI, and visualized utilizing Jalview application in order to investigate nucleotide sequence *Mal d 1* gene family variability.

2.6 Phylogenetic Analysis

Neighbor-joining tree has been constructed using Mega v5 with 500 bootstrap replications and p-distance for amino acid sequences of *Mal d 1* alleles revealed in the present study as well as all deposited in the NCBI database. For this all non-redundant amino acid sequences of *Mal d 1* alleles were selected from the NCBI database.

2.7 Inferred Biomolecular Interaction (IBIS, NCBI)

For the prediction of potential biomolecular interactions an Inferred Biomolecular Interaction Server (IBIS, NCBI) has been used. For a given protein sequence IBIS reports physical interactions observed in the experimentally determined structures for the particular protein and infers/predicts interacting partners and binding sites by homology.

3. Results

*Mal d 1* is coded by a large gene family of 18 members mapped on three linkage groups of the apple genome i.e. LG6 (*Mal d 1.05*), LG13 (*Mal d 1.01*), and LG16 (*Mal d 1.02, Mal d 1.04, Mal d 1.06A, Mal d 1.06B, Mal d 1.06C*) (Atkinson et al., 1996; Gao et al., 2005a, 2005b, 2005c). Not all of these members are likely to be involved in allergenicity since only a limited number of different *Mal d 1* proteins and mRNAs have been traced back in apple fruit so far (Helsper et al., 2002; Puehringer et al., 2003; Beuning et al., 2004).

Studies done by Son et al. (1999) showed that allergenic differences between apple cultivars are mainly related to the expression levels of *Mal d 1* and not to the presence of different isoformes, whereas Gao et al. (2008) stated that differences in allergenicity are associated with the allelic composition of two specific genes (*Mal d 1.04* and *Mal d 1.06 A*).


Here we identified alleles of the *Mal d 1.01*, *Mal d 1.02*, *Mal d 1.04*, *Mal d 1.06A*, *Mal d 1.06B*, and *Mal d 1.06C* in all investigated cultivars (Table 2).

Table 2. *Mal d 1* alleles identified in following Polish contemporary and historical apple cultivars ‘Kantowka Gdanska (‘Danziger Kantapfel’), ‘Kosztela’, ‘Ligol’, and ‘Redkroft’. Accession numbers for nucleotide and deduced protein sequences, as well as genomic positions are provided.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Gene accession number; deduced protein accession number</th>
<th>Investigated cultivar</th>
<th>Other cultivars</th>
<th>Genomic position</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mal d 1.01</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mald1.0105.05</td>
<td>JF682500;AEE38275</td>
<td>‘Danziger Kantapfel’</td>
<td>-</td>
<td>chr13:14587571..14589168</td>
</tr>
<tr>
<td>Mald1.0105.04</td>
<td>JF682501;AEE38276</td>
<td>‘Danziger Kantapfel’</td>
<td>Red Delicious (AY827643), Fuji (AY827644)</td>
<td>chr13:14587571..14589168</td>
</tr>
<tr>
<td>Mald1.0105.07</td>
<td>JF682503;AEE38278</td>
<td>‘Kosztela’</td>
<td>-</td>
<td>chr13:14587571..14589168</td>
</tr>
<tr>
<td>Mald1.0105.08</td>
<td>JF682505;AEE38280</td>
<td>‘Kosztela’</td>
<td>-</td>
<td>chr13:14587571..14589168</td>
</tr>
<tr>
<td>Mald1.0105.02</td>
<td>JF682507;AEE38282</td>
<td>‘Ligol’</td>
<td>Discovery (AY827638), Fiesta (AY879238), GD (AY827633), Redkroft (JF682511)</td>
<td>chr13:14587571..14589168</td>
</tr>
<tr>
<td>Mald1.0105.01</td>
<td>JF682508;AEE38283</td>
<td>‘Redkroft’</td>
<td>Golden Delicious (AY827639), Jonathan (AY827640), Fuji (AY827641)</td>
<td>chr13:14587571..14589168</td>
</tr>
<tr>
<td>Mald1.0105.10</td>
<td>JF682509;AEE38284</td>
<td>‘Redkroft’</td>
<td>-</td>
<td>chr13:14587571..14589168</td>
</tr>
<tr>
<td><em>Mal d 1.02</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mald1.0210</td>
<td>JF507716;AEE38452</td>
<td>‘Danziger Kantapfel’</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mald1.0211</td>
<td>JF507717;AEE38453</td>
<td>‘Danziger Kantapfel’</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mal d 1.01</td>
<td>JF507718;AEF38454</td>
<td>‘Kosztela’</td>
<td>-</td>
<td>chr16:10767391..10782496</td>
</tr>
<tr>
<td>Mal d 1.02</td>
<td>JF507719;AEF38455</td>
<td>‘Kosztela’</td>
<td>Prima (AY789240), Fiesta (AY789241)</td>
<td>chr16:10767391..10782496</td>
</tr>
<tr>
<td>Mal d 1.01</td>
<td>JF507721;AEF38457</td>
<td>‘Ligol’</td>
<td>-</td>
<td>chr16:10767391..10782496</td>
</tr>
<tr>
<td>Mal d 1.01</td>
<td>JF507723;AEF38459</td>
<td>‘Ligol’</td>
<td>-</td>
<td>chr16:10767391..10782496</td>
</tr>
<tr>
<td>Mal d 1.01</td>
<td>JF507724;AEF38460</td>
<td>‘Redkroft’</td>
<td>-</td>
<td>chr16:10767391..10782496</td>
</tr>
</tbody>
</table>

All Mal d 1 sequences were aligned to the *M. x domestica* genome v1.0 (http://www.rosaceae.org/species/malus/malus_x_domestica genome_v1.0) in order to define their genomic positions. One-locus representing genes were Mal d 1.01, 1.02 and 1.04. Mal d 1.01 which have been localized on chromosome 13 encompassing the region of 14587571-14589168 nt, Mal d 1.02 – positioned on chromosome 16 occupying locus ranging from 10767391 till 10782496 nt, while Mal d 1.04 – on chromosome 16 in the range of 11352896-11357492 nt. Mal d 1.06 allelic variants were mapped to three different positions on the chromosome 16, namely 10738273-10741966 nt (Mal d 1.06A), 10806399-10818287 nt (Mal d 1.06B), and 11433700-11436518 nt (Mal d 1.06C) (Table 2).
3.2 Allele Verification

In order to validate allelic variability of potential *Mal d 1* variants identified using Sanger approach, the 454 sequencing has been deployed using primers designed to encompass single nucleotide polymorphisms (SNPs) (Supplementary Table 1).

All together we identified seven variants of *Mal d 1.01* (three new), three variants of *Mal d 1.02* (two new), four variants of *Mal d 1.04* (three new) and pseudoallele *Mal d 1.04ps2*, seven variants of *Mal d 1.06A* (three new), six variants of *Mal d 1.06B* (four new), three variants of *Mal d 1.06C* (two new) (Table 2, Figures 1a, 1b, 1c, 1d, 1e, and 1f).

(a)

(b)
Figure 1. Multiple nucleotide sequence alignment of a) Mal d 1.01, b) Mal d 1.02, c) Mal d 1.04, d) Mal d 1.06A, e) Mal d 1.06B, and f) Mal d 1.06C alleles

3.3 Nucleotide Sequence Diversity of Mal d 1 Gene Family

Multiple sequence alignments of newly identified alleles have been performed in order to assess phylogenetic distance between Mal d 1 allelic variants and antigens derived from other food sources.

Some molecular analyses (Gao et al., 2005a, 2008) revealed that Mal d 1 gene family can be subdivided into two major categories: genes with and without the intron. We found that intron containing variants of Mal d 1 were all alleles of Mal d 1.01 and 1.02. Mal d 1.06 from A to C varied with respect to the number and positions of introns. Among analyzed sequences intronless appeared to be Mal d 1.04 alleles. We found 7 nucleotide polymorphisms among Mal d 1.01 aligned sequences. As many as 6 and 5 SNPs were in Mal d 1.02 and Mal d 1.04. The most variable was Mal d 1.06A in which 29 nucleotide polymorphisms were found among 9 alleles, while in Mal d 1.06B and Mal d 1.06C – 21 and 27, respectively (Figures 1a, 1b, 1c, 1d, 1e, and 1f).

3.4 Deduced Amino Acid Sequence Diversity of Mal d 1 Gene Family

Our studies revealed that Mal d 1.01 is diverged at the genomic level but showed to be conserved at the protein level with two protein isoforms deduced from seven allelic variants (Table 3). The other proteins encoded by this family were more variable, with three isoforms deduced from Mal d 1.02, Mal d 1.04, Mal d 1.06A, and four deduced from Mal d 1.06C (Table 3).

Table 3. Number of identified allelic variants and deduced protein isoforms of Mal d 1 gene family, identified in Polish historical and contemporary cultivars

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alleles</th>
<th>Protein isoforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mal d 1.01</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Mal d 1.02</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Mal d 1.04</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Mal d 1.06A</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Mal d 1.06B</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Mal d 1.06C</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

3.5 Phylogenetic Analysis

Neighbor-joining tree of both Mal d 1 apple allergens as well as allergens derived from other food sources resulted in five well distinguished clusters. The most phylogenetically distant group included allergens derived from celery, carrot, soybean and potato. They were not related to any of the Mal d 1 alleles. Another group consisted of Mal d 1 alleles grouped with allergens from Fragaria, Pyrus, and Prunus avium. Interestingly alleles of Mal d 1.04 were clustered with allergen from apricot. Mal d 1.06 A, B, and C and Mal d 1.04B created another distant group, not related to other taxa (Figure 2).
Figure 2. Phylogenetic analysis of Mal d 1 alleles identified in ‘Kantowka Gdanska (‘Danziger Kantapfel’), ‘Kosztela’, ‘Ligol’, and ‘Redkroft’. Mal d 1 alleles originated from other apple cultivars, as well as allergens from Alnus glutinosa (Aln g 1), Apium graveolens (Api g 1), Betula pendula (1FM4A), Betula verrucosa (Bet v 1), Corylus avellana (Cor a 1), Daucus carota (Dau c 1), Fragaria x ananassa (Fra a 1), Glycine max (Gly m Bd 30K), Prunus armeniaca (Pru ar 1), Prunus avium (Pru av 1), Pyrus communis (Pyr c 1), and Solanum tuberosum (Sola t 2)
3.6 Prediction of the IgE Binding Sites to the Amino Acid Residues of Identified Mal d 1 Alleles

Apple allergen cross-reactivity results from Bet v 1 specific IgE binding capacity to homologous epitopes of both Bet v 1 and Mal d 1 (Holm et al., 2011). Detailed examination of regions of sequence which serve as epitopes is crucial in research targeted at developing hypoallergenic cvs. It was observed that these regions, i.e. hydrophobic binding sites and glycine rich P loop, are quite conserved between Mal d 1 proteins and Bet v 1 from birch and Gly m 4 from soybean (Jenkins et al., 2005). Glycine-rich loop has been demonstrated to serve as IgE epitope in PR-10 proteins with allergenicity (Mirza et al., 2000).

We aligned the conserved binding sites of IgE antibodies to amino acid residues of the known three dimensional structure of the 1FM4 antigen from Betula species. Using Inferred Biomolecular Interactions NCBI server it was possible to define amino acid residues interacting with a Fab fragment of monoclonal IgG Antibody and the major allergen from birch pollen Bet V 1 (accession number 1FM4) (Figure 3).

It has been already revealed that crucial regions of the binding to the light chain of the IgG antibody kappa consists of following amino acid residues: N 47, G 48, G 49, P 50, G 51, D 72, and H 76, while those interacting with the heavy chain: E 42, N 43, I 44, E 45, G 46, N 47, P 50, G 51, T 52, I 53, R 70, D 72, E 87, and K97 (Ma et al., 2006). Glutamic acid 46 is situated in the center of the epitope. The epitope must be classified as discontinuous. However, residues 42–52 (including the alleged “P-loop like” region 46–51 of the Bet v 1 sequence) constitute 80% of the contact surface (Ma et al., 2006).

In our approach we compared Mal d 1 amino acid sequences of newly identified alleles and those from the NCBI database from celery, hazelnut, birch, apricot, strawberry, alder, potato, sour cherry, pear, and soybean, to the conserved regions of Bet v 1 and Gly m 4 binding sites. The analysis revealed the potential amino acid residues that can serve as interaction regions during sensitization reaction between the antigen and the antibody (Figure 3). This kind of analysis has been also anticipated to indicate potential cross-interacting allergens derived from apple and other food sources. Additionally IBIS analysis (Inferred Biomolecular Interaction Server) revealed Mal d 1 residues putatively reacting with light and heavy chains of immunoglobulin antibody. These residues are marked with arrows and correspond to glycine rich P loop of the epitope region (Figure 3).
Figure 3. IBIS analysis results for Mal d 1 alleles identified in ‘Kantowka Gdanska (‘Danziger Kantapfel’), ‘Kosztela’, ‘Ligol’, and ‘Redkroft’. Mal d 1 alleles originated from other apple cultivars, as well as allergens from Alnus glutinosa (Aln g 1), Apium graveolens (Api g 1), Betula pendula (1FM4A), Betula verrucosa (Bet v 1), Corylus avellana (Cor a 1), Daucus carota (Dau c 1), Fragaria x ananassa (Fra a 1), Glycine max (Gly m Bd 30K), Prunus armeniaca (Pru ar 1), Prunus avium (Pru av 1), Pyrus communis (Pyr c 1), and Solanum tuberosum (Sola t 2). Mal d 1 residues putatively reacting with light and heavy chains of immunoglobuline antibody are marked with arrows. The P-loop region and residues of the epitope region which serve as potential interacting partner for light and heavy kappa chains of the immunoglobuline are indicated by the red box.
The most variable positions of the interacting regions revealed for the heavy chain of the IgG antibody included proline P50 present in the Mal d 1.02 alleles, and all other antigens from food sources analyzed excluding carrot (alanine A50), potato (threonine T50), soybean (serine S50), and apricot (valine V50). In rest of Mal d 1 alleles this position was substituted by valine (V). The most conserved residues were threonine (T52) and isoleucine (I53), only variable in celery, carrot, and potato (Figure 3).

Further the phylogenetic analysis of the amino acid region limited to the residues which serve as interaction partners with the antibody showed that Mal d 1.06A, B, and C isoforms are much more related to soybean allergen contrary to Mal d 1.01, 1.02 and 1.04 putative epitopes which are grouped together with the sweet cherry and Bet v 1 epitopes (Figure 4).
It can denote that *Mal d 106 A*, B and C putative isofoms are much more probable to cross-react with allergen from soybean (Gly m 4) than *Mal d 1.01*, 1.02, and 1.04 in which epitopes are predicted to cross-react with Bet v 1 and *Pru a 1* from birch and sweet cherry, respectively.

4. Discussion

The major apple allergen *Mal d 1* is homologous to the major birch pollen allergen Bet v 1. Primary sensitization occurs via Bet v 1, resulting in cross-reactivity of IgE to *Mal d 1* (Fernández-Rivas et al., 2006).

The apple allergen *Mal d 1* and the major birch pollen allergen, Bet v 1 share allergenic epitopes leading to IgE cross-reactivities (Holm et al., 2011). Especially during the birch pollen season, an increase in clinical reactions to apples occurs (Skamstrup-Hansen et al., 2001). Determination of the allelic *Mal d 1* allergen constitution together with the medical assessment can serve as a first step to elucidate the basis of differences in allergenicity among cultivars and to develop breeding schemes for new hypo-allergenic cultivars. In addition, this know-how might be used by apple growers to develop novel cultivars with decreased allergenicity. Further fruit with low allergenic potential might be tolerated by patients with mild apple allergy.

Here we aimed at investigation of the major apple allergen *Mal d 1* allelic variability in Polish contemporary and historical apple cultivars. *Mal d 1.01* is known as a single locus gene with maximum two alleles present in a cultivar (Gao et al., 2005a). In order to avoid sequencing and cloning mistakes listed by Gao et al. (2008) we cloned presented variants in two independent experiments. Specific primers encompassing polymorphic regions have been designed in order to identify allelic variants. Only sequences validated by both Sanger and 454 sequencing technologies were taken into account.

Investigation of the following Polish historical and contemporary apple cultivars ‘Kantowka Gdanska’ (‘Danziger Kantapfel’), ‘Kosztela’, ‘Ligol’, ‘Redkroft’, and ‘Odra’ revealed three new variants of *Mal d 1.0105*, two new variants of *Mal d 1.02*, three new alleles of *Mal d 1.04* and pseudoallele *Mal d 1.04ps2*, three new alleles of *Mal d 1.06A*, four new alleles of *Mal d 1.06B*, and two new alleles of *Mal d 1.06C*.

We compared newly identified alleles to known apple allergens as well as those originated from other plant taxa. The most distant phylogenetic group included allergens derived from celery, carrot, soybean and potato. They were not related to any of the *Mal d 1* alleles. *Mal d 1.0105* alleles appeared to be closely related to allergens from strawberry, pear, and sour cherry. Interestingly alleles of *Mal d 1.04* were clustered with allergen from apricot, while *Mal d 1.06 A*, B, and C and *Mal d 1.04B* created another distant group and were not related to any of the allergens derived from other taxon. These data may point to variability in cross-reactivity between allergens originated from *Malus x domestica* and other plant taxons.

Although cross-reactivity is mostly determined by 3D protein structure and conformational epitopes, the high percentage of identity between analyzed antigens can allow predicting possible IgE cross-reactions. IgE antibodies specific for the major birch pollen allergen Bet v 1, have been shown to cross-react with homologous proteins identified in different fruits and vegetables, such as apple (Mal d 1), sweet cherry (Pru av 1), pear (Pyr c 1), hazelnut (Cor a 1), celery (Api g 1), carrot (Dau c 1), soybean (Gly m 4), peanut (Ara h 8), jackfruit, kiwi (Act d 8), strawberry (Fra a 1), apricot (Pru ar 1), potato (Sola t 1), mungbean (Vig r 1), alder (Aln g 1), hornbeam (Car b 1), chestnut (Cas s 1), hazelnut (Cor a 1), and pollen of white oak (Que a 1). These food allergens can also activate Bet v 1–specific T cells to proliferate and produce cytokines, sometimes even resulting in T cell–mediated late-phase responses in target organs, such as flare-ups of atopic eczema in the skin (Geroldinger-Simic et al., 2011).

The tertiary structure of Bet v 1 contains three surface patches that have conserved-amino acid sequences and structures with proteins from a variety of food including cherry, apple, hazelnut, peach, carrot, celery, and soya (Vieths et al., 2002). Among them conserved-IgE cross reactive region is the P-loop located between b-strands 2 and 3 of Bet v 1 and its homologs (Mizra et al., 2000; Neudecker et al., 2003). Elicitation of clinical allergy symptoms appears to be dependent on IgE binding to these structures, which is in direct contrast to that observed for other food allergens with linear IgE binding epitopes such as plant allergens from peanut (Ara h 1, Ara h 2, Ara h 3), soya (G2 Glycinin, P34/Gly m Bd 30K), wheat (ω-5 Gliadin), pollen (Jun a 1, Par j 1, Par j 2), nuts (Jug r 1, Ana o 1), and animal allergens from egg (Ovalbumin, Ovomucoid), shrimp (Tropomyosin), and milk (αs2-Casein) (Bannon & Ogawa, 2006).

We predicted potential binding sites of apple allergens to the IgE antibody. The most variable positions of the heavy chain IgG antibody interacting regions included proline P50 present in the *Mal d 1.02* alleles, Bet v 1, and *Pru av 1*. In the rest of *Mal d 1* alleles this position was substituted by valine (V). In the carrot, potato, soybean, and apricot there were A50, T50, S50, and V50, respectively. The most conserved residues were threonine (T52).
and isoleucine (I53), variable only in celery, carrot, and potato.

The 3-D structure prediction reveals which amino acids are surface exposed and accessible for interactions with Abs. Combining it with primary sequence alignment analysis, we were able to indicate the conserved surface structures of homologous allergens derived from different species.

Glycine-rich loop, GXGXXGXXK has been previously reported to be highly conserved among all the representative members of Bet v 1-like proteins in different species (Spangfort et al., 1997), and our studies on identified Mal d 1 isoforms, comprising the GDGGVGTKK residues fully support this. Only aspartic acid (D47) has been substituted by asparagine (N) in Mal d 1.0105.05 (AEE38275), and valine (V50) – by proline (P) in Mal d 1.0105.05 (AEE38275), Mal d 1.0210 (AEE38452) and Mal d 1.0211 (AEE38453), giving the same P-loop region as in the Bet v 1, Cor a 1, 1FM4A, Pyr c 1, and Dau c 1. Pagliarani et al. (2012) reported few substitutions in this domain among the predicted Mal d 1 proteins: a lysine (K) replaced by a glutamine (Q) in Mal d 1.08 and by a methionine (M) in Mal d 1.09; the third glycine replaced by glutamic acid (E) in Mal d 1.11A and Mal d 1.11B and by arginine in Mal d 1.12. The information on immunological properties of the various isoforms of both the Bet v 1 and Mal d 1 families, and the ability of these different proteins to induce allergic responses is scarce. Besides it has been shown that a high sequence similarity between proteins increases the chance to be recognized by the same antibody, a single amino acid change in a key position may drastically influence the extent of allergenicity by increasing or decreasing the ability to bind that antibody. Wagner et al. (2008) showed that a few amino acid changes (from four to nine changes) on the surface of three Bet v 1 isoforms caused a difference in IgE induction. The S113C change of Bet v 1.04 with respect to Bet v 1.01 has previously been identified as being important for the ability of Bet v 1.04 to form aggregates and to create a type of protection against IgE binding (Zaborsky et al., 2010). Pagliarani et al. (2012) considered Mal d 1.11A/B and Mal d 1.03C/I/H as the best candidates for similar IgE binding protection because of the presence of C113. In our studies among all identified Mal d 1 isoforms this position consisted either isoleucine (I) as in Fra a 1, Api g 1, Pyr c 1, and Dau c 1 or leucine (L) as in Mal d 1.0407.02, Mal d 1.0404.05, Bet v 1, Pru av 1, Cor a 1, 1FM4A.

Further a previous crystallographic study of the Bet v 1–antibody complex classified several residues as important for the antigenic surface, including the positions E42, E45, T52, R70, D72, H76 and K97. In particular, position 45 seems to be crucial because it is located in the core of the binding pocket of Bet v 1.01 and fits into the groove on the antibody surface (Ghosh & Gupta-Bhattacharya, 2008). This position has been shown to be conserved in apple and birch, what corresponds to our findings with the exceptions for Gly m Bd 30K where E45 was substituted by arginine (A), Api g 1 and Dau c 1 – by lysine (K), and Sola t 2 – by valine (V). In another studies E45 has been found to be substituted by serine in Mal d 1.11A and B, which both are considered as a putative hypoallergenic Mal d 1 isoforms (Pagliarani et al., 2012).

Thanks to multiple sequences alignment limited to residues which serve as interaction partners for the antibody we were able to assess more precisely the phylogenetic distance between analyzed antigens with respect to these putative epitope sequence residues. We found that Mal d 1.06A, B, and C isoforms are much more related to soybean allergen contrary to Mal d 1.01, 1.02 and 1.04 putative epitopes which are grouped together with the sweet cherry and Bet v 1 epitopes. It can denote that Mal d 1.06A, B, and C putative isoforms are much more probable to cross-react with allergen from soybean (Gly M 4) than Mal d 1.01, 1.02, and 1.04 in which epitopes are predicted to cross-react with bet v 1 and Pru a 1 from birch and sweet cherry, respectively.

Our approach aimed at the identification of Mal d 1 alleles in historical and contemporary Polish cultivars and the investigation of their variability allowed finding these variants which possess the potential for the interaction with allergens of great importance, derived from birch and food sources as soybean. Particular allelic variants are supposed to interact with specific isoforms what gives a possibility to predict possible interactions between antigens and focus breeding programs towards development new apple cultivars of lower allergenicity. However these results need further in vivo experimental investigations of cross-reactivity between Mal d 1 isoforms and Bet v 1 allergen to assess the level of apple cultivar hypo/hyperallergenicity as well as clinical assessments. This is especially important since it is well known that IgE antibodies specific for Mal d 1 allergens can cross-react with antigens from other food derived products.

**Acknowledgements**

These studies were supported by the Polish Ministry of Science and Higher Education (grant No. NN 310 0252 33).
Competing Interests
The authors have no competing interests.

Authors’ Contributions
KLB and AAP conceptualized the project; KLB performed laboratory work; NK performed bioinformatics’ analysis; MF participated in molecular work design; KLB, AAP, NK, PW, and MF were involved in manuscript drafting and writing.

References


Ma, Y., Gadermaier, G., Bohle, B., Bolhaar, S., Knulst, A., Markovic-Housley, Z., ... Ferreira, F. (2006). Mutational analysis of amino acid positions crucial for IgE-binding epitopes of the major apple (Malus domestica) allergen, Mal d 1. *International Archives of Allergy and Immunology, 139*(1), 53-62. http://dx.doi.org/10.1159/000089756


Neudecker, P., Lehmann, K., Nerkamp, J., Haase, T., Wangorsch, A., Fotisch, K., ... Scheurer, S. (2003). Mutational epitope analysis of Pru av 1 and Api g 1, the major allergens of cherry (Prunus avium) and celery (Apium graveolens), correlating IgE reactivity with three-dimensional structure. *The Biochemical Journal, 376*(Pt 1), 97-107. http://dx.doi.org/10.1042/BJ20031057


Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).