Food Chemistry: Molecular Sciences 4 (2022) 100111

Contents lists available at ScienceDirect

Food Chemistry: Molecular Sciences

journal homepage: www.sciencedirect.com/journal/food-chemistry-molecular-sciences

Identification of allergenomic signatures in allergic and well-tolerated apple genotypes using LC-MS/MS

Soraya Chebib [a], Chen Meng [b], Christina Ludwig [b], Karl-Christian Bergmann [c], Sylvia Becker [d], Werner Dierend [e], Wilfried Schwab [a,*]

[a] Biotechnology of Natural Products, Technical University Munich, Liesel-Beckmann-Str. 1, 85354 Freising, Germany
[b] Bavarian Center for Biomolecular Mass Spectrometry (BayBioMS), Technical University Munich, Gregor-Mendel-Straße 4, 85354 Freising, Germany
[c] Allergy-Centre-Charité, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Germany
[d] European Centre Allergy Research Foundation, Robert-Koch-Platz 7, 10115 Berlin, Germany
[e] Faculty of Agricultural Science and Landscape Architecture, Fruit Science, University of Applied Sciences Osnabrück, Oldenburger Landstr. 24, 49090 Osnabrück, Germany

ARTICLE INFO

Keywords:
Malus domestica
Mal d proteins
Proteomics
LC-MS/MS
Isoforms
Apple breeding

ABSTRACT

The apple fruit (*Malus domestica* L. Borkh) is one of the most popular fruits worldwide. Beyond their beneficial properties, apples contain proteins that trigger allergic reactions in susceptible consumers. Mal d1 to d4 are allergens present in a variety of different isoforms in apples. In this study, we used proteomics to quantify all four Mal d proteins in 52 apple genotypes with varying allergenic potentials. A total of 195, 17, 14, and 18 peptides were found to be related to Mal d1, d2, d3, and d4 proteins, respectively of which 25 different Mal d proteins could be unambiguously identified. The allergenic potential of the Mal d isoforms was characterized by comparing the isoform abundance with the allergenic score of genotypes from oral challenge tests. The detected Mal d peptides presumably have different IgE binding properties and could be used as potential molecular markers to discriminate between hypoallergenic and hyperallergenic cultivars.

1. Introduction

The apple (*Malus domestica* L. Borkh), a fruit of the rose family (Rosaceae), is cultivated and consumed worldwide. In addition to multiple health benefits of its phytochemical constituents, apples are the main cause of fruit allergies in Northern and Central Europe and in North America (Burney et al., 2014; Fernández-Rivas et al., 2006; Matricardi et al., 2016). Allergic individuals sensitized to birch pollen often also show allergic reactions to apple fruits, due to the cross-reaction between the main birch pollen allergen Bet v1 and its structural homologue Mal d1, the main apple allergen (Holm et al., 2001). Both are members of the pathogenesis-related protein 10 (PR-10) family, indicating an important role in plant defense. Mal d1, a heat labile protein sensitive to pepsin digestion, causes relatively mild and local symptoms including itching, tingling, or swelling of lips, tongue and throat, known as oral allergy syndrome (OAS). In southern Europe apple allergy is not related to birch pollen allergy and occurring symptoms are more severe. This type of allergy can be linked to the non-specific lipid transfer protein (nsLTP) Mal d3, that is resistant to high temperature and enzymatic digestion (Sánchez-Monge et al., 1999). Other allergens that have been reported and characterized in apple are the profilin Mal d2 and the thaumatin-like protein Mal d4 (Breiteneder & Radauer, 2004).

The extent of allergic reactions of allergic sufferers varies with the consumption of different apple varieties (Bolhaar et al., 2005a). The cultivar Santana (SAN) has been extensively studied and has been shown to be better tolerated by most allergic sufferers, unlike the Golden Delicious (GD) variety, and is considered suitable for individuals with OAS (Bolhaar et al., 2005a). Decisive factors leading to variations in the

Abbreviations: ELISA, Enzyme-Linked Immunosorbent Assay; LC-MS/MS, Liquid Chromatography tandem Mass Spectrometry; PR-10, Pathogenesis-Related protein 10; OAS, Oral Allergy Syndrome; nsLTP, non-specific Lipid Transfer Protein; SAN, Santana; GD, Golden Delicious; IgE, Immunoglobulin E; PSM, Peptide-Spectrum Match.

* Corresponding author.
E-mail addresses: soraya.chebib@tum.de (S. Chebib), chen.meng@tum.de (C. Meng), tina.ludwig@tum.de (C. Ludwig), Sylvia.Becker@ecarf.org (S. Becker), W.Dierend@hs-osnabrueck.de (W. Dierend), wilfried.schwab@tum.de (W. Schwab).

https://doi.org/10.1016/j.fochms.2022.100111
Received 13 August 2021; Received in revised form 12 April 2022; Accepted 2 May 2022
Available online 9 May 2022

allergenic potential of apple cultivars are the total allergen content, mainly related to Mal d1 and different expression levels of its isoforms (Son et al., 1999). In this context, the term isoform includes paralogs and alleles rather than splice forms. For Mal d1, over 100 isoforms exist, that have been identified using biochemical sequencing techniques (Son et al., 1999; Pagliarani et al., 2013; Gao et al., 2005; Beuning et al., 2004; Romer et al., 2020). Isoforms show a high sequence identity and thus structural similarity and therefore share physiochemical properties. The allergenicity of a protein is primary defined by the occurrence of IgE binding epitopes, which can be either sequential or conformational. In contrast to sequential epitopes, which are linear sequences of amino acids, conformational epitopes are discontinuous residues within the sequence that are brought together in the folded structure (Sanchez-Trincado et al., 2017; Potocnakova et al., 2016). Sequential and conformational epitopes of Mal d1 were previously studied using single point mutation experiments (Ma et al., 2006; Uehara et al., 2001). In contrast, little is known about proteomic profiles of apple genotypes, the diversity of Mal d isoforms in the genotypes, their expression on protein level and their allergenicity.

The availability of the fully mapped apple genome provided the basis for transcriptomic and proteomic studies. Preliminary proteomic studies focused on significant changes of expressed proteins during maturation and fruit ripening to gain deeper knowledge of metabolic processes (Shi et al., 2014). A proteomic approach was used to map the proteome of apple fruit, juice and cider and allowed the identification of differentially expressed proteins including Mal d3 (Lerma-García et al., 2019). Recently, an exploratory proteomic study was performed to detect Mal d1 isoforms in the two apple cultivars SAN and GD, revealing different isoform patterns (Romer et al., 2020). The authors presented a combinational approach in which Mal d1 content was also determined by ELISA, the polyphenol profile was analysed by LC-MS, and human studies were performed with oral challenge tests to assess the allergenic potential of different apple genotypes. It was shown that the allergenicity of apple genotypes does not always correlate with the Mal d1 content and that other criteria such as the flavan-3-ol content and the isoform distribution can be potential influencing factors. However, as the expression of Mal d-isoforms and their allergenic potential are still less well characterized further research is needed (Son et al., 1999; Puehringer, 2003; Gao et al., 2005; Bolhaar et al., 2005b; Paris et al., 2017).

In this study, we determined the isoform composition of the Mal d1 proteins in different apple genotypes using a combination of bioinformatic and proteomic tools. All genotypes analyzed by proteomics were previously tested in a human study among patients with a history of allergic reactions to apples and sensitized to Bet v1. Thus, as a second goal we sought correlations between the occurrence of isoforms in specific genotypes and the tolerability of apple cultivars. Although our focus was set on Mal d1, the frequency of Mal d2 to d4 was also investigated. The results showed that proteomics represents a powerful tool for the identification of allergen isoforms. The detected peptides can be affiliated to varying IgE-binding properties and are potentially important for the breeding of hypoallergenic apple cultivars.

2. Methods

2.1. Fruit materials

The University of Applied Sciences Osnabrück in cooperation with the organization "Züchtungsinitiative Niederelbe GmbH & Co. KG" provided apple fruits. Fruits were harvested between August and October in both 2018 and 2019. All genotypes used here are the same as recently described (Romer et al., 2020), except for one genotype from the harvest year 2018 (p17) and five genotypes from the harvest year 2019 (p168, p185, p78, p211, p92, Santana (SAN)), that were additionally used for analyses. The genotype p29 (2017, 2018) is the cultivar Jonagold. The genotype p186 from all harvest years was renamed to p211. Further treatment and processing of the apple samples was carried out according to Romer et al. (2020).

2.2. Protein extraction

Apples were ground to a fine powder in liquid nitrogen. Protein sample preparation and extraction was carried out according to Romer et al. (2020).

2.3. Mal d1 determination

Recombinant Mal d1.02 was produced and purified as recently described by Romer et al. (2020). Mal d1 content was determined by indirect competitive ELISA using a previously published method (Romer et al., 2020) with one modification, as a carbonate/bicarbonate buffer (30 mM Na_2CO_3, 70 mM $NaHCO_3$, pH 9.6) was used as coating buffer.

2.4. Human study

The human study was conducted as previously described with slight modifications for the genotypes of the harvest year 2019 (Romer et al., 2020). Ethical approval was obtained from the Charité ethic commission (No. EA1/311/19). Ten to twenty-one adult patients with clinical allergy to birch pollen and apple were examined outside the birch pollen season. The study participants comprised a group of 20 women and one man between the age of 25 and 60 years. The patients were subjected to an oral provocation test with fresh apples. Fruits were stored up to 5 months at 2.5 °C and under controlled atmosphere, including ultralow oxygen storage conditions (1.5% O_2; 1% CO_2). Apples were given to patients in increasing doses of approximately 20 g, 40 g, 80 g and the rest of the apple. Symptoms like itching and mild swelling of mouth and throat were recorded 20 to 30 min after each dose. A four-point scale was used for scoring the symptoms from zero to three, where zero describes the absence of symptoms and three the occurrence of severe symptoms. For each patient, all symptom scores recorded after eating the remaining apple were summed. Finally, the mean value was calculated from the total number of scores and used to classify the genotypes in terms of their allergenic potential. The oral provocation test was discontinued when clinical symptoms with a severity of two occurred. The mean symptom score was used to subdivide the different genotypes with regard to statistical analyses.

2.5. Protein digest fractionation and LC-MS/MS

Tryptic in-gel digestion and LC-MS/MS measurements were performed as described previously (Romer et al., 2020).

2.6. Protein identification, quantification and statistical analysis

All the raw files acquired by mass spectrometry were loaded into the MaxQuant software (version 1.6.3.4; Cox & Mann, 2008) and searched against an apple (*Malus domestica*) proteome database downloaded from National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) containing 52,099 protein sequences (Romer et al., 2020). The peptide search was performed using carbamidomethyl cysteine as fixed modification, oxidation of methionine and acetylation of protein N-terminus as variable modifications. Trypsin was specified as proteolytic enzyme and up to two missed cleavages were allowed. Protein identifications were filtered to 0.01% false-discovery rate on peptide-spectrum match (PSM) and protein-level. The peptide intensity was used to represent the relative abundance of peptides. T-test was applied to compare the abundance of proteins between two groups. Pearson correlation were utilized for the correlation analysis. The hierarchical cluster analysis was performed using Ward's method; the number six, a value close to the lowest detected intensity, replaced missing values in the LC-MS/MS dataset. All statistical analysis were

performed using R (v3.6.3). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (http://www.ebi.ac.uk/pride) partner repository with the dataset identifier PXD027753 (Project Name: Identification of allergenomic signatures in allergic and well-tolerated apple genotypes using LC-MS/MS).

2.7. Phylogenetic analysis of Mal d proteins

The phylogenetic tree was generated with Geneious Prime (v. 2021.2.2) (https://www.geneious.com) and annotated using iTOL (v. 6; Letunic and Bork (2019)). The tree was built by the Neighbor-Joining method with no outgroup. The genetic distance model was Jukes-Cantor. For better visibility, the branches were transformed so that no scale bar could be displayed, but the sequence similarities are shown in Supplementary Table S1. The subdivision into Mal d1 groups was carried out according to NCBI database entries. For those not assigned to a group according to the NCBI database, the classification was based on their phylogenetic relationships.

3. Results

3.1. Definition of genotype groups

The apple genotypes used in this study originate from segregation populations generated by the Osnabrück University of Applied Sciences in cooperation with the privately funded "Züchtungsinitiative Niederelbe GmbH & Co. KG". Diverse crossing combinations produced the breeding lines. Established cultivars as well as pre-selected breeding lines were used as parents, reflecting a profile of high quality apples. Honeycrisp, Braeburn, Gala, Dalinbel, Santana, Pinova, Gloster, Topaz, Golden Delicious, Rubinette, Elstar, Delbarestivale, Retina, Rubens, Fuji, and Nicoter were used as parent cultivars. Thus, the genotypes should express a variety of different Mal d1 isoforms. Allergic sufferers assigned the analysed genotypes of the consecutive harvest years 2017, 2018 and 2019 to three groups of apple genotypes according to their tolerability. By averaging all symptom values obtained for a genotype in the human study, a symptom score was calculated for each genotype (Supplementary Fig. S1). Group I comprises well-tolerated genotypes with a symptom score below 1.4, group II includes moderately tolerated genotypes with a symptom score between 1.4 and 1.7, and group III contains allergenic genotypes with symptom scores above 1.7 (Supplementary Fig. S2).

3.2. Analysis of the Mal d allergens by mass spectrometry based proteomics

To enable comprehensive characterization of Mal d proteins using proteomics, we extracted all protein sequences of *Malus domestica* L. Borkh from the NCBI database and used them for protein identification. The resulting protein database contained 113 Mal d1 isoforms, 10 Mal d2 proteins, 9 Mal d3 proteins and 15 Mal d4 proteins (Fig. 1a). NCBI database entries combined with phylogenetic analysis subdivided all Mal d1 proteins into 12 groups. Within the subdivided groups, pairwise identity of protein sequence ranged between 93 and 99 % (Supplementary Table S1).

We used a mass spectrometry-based proteomics approach to quantify proteins in our 52 apple genotypes. A total of 50,941 tryptic peptides (Supplementary Table S2) were found, among which 195, 17, 14, and 18 peptides were related to Mal d1, Mal d2, Mal d3, and Mal d4 proteins, respectively (Fig. 1b). Twenty-five different isoforms of Mal d proteins could be unambiguously identified (21 Mal d1-, 2 Mal d3-, and 2 Mal d4-proteins; Table 1) by 45 isoform-specific peptides, which occurred only in a single Mal d isoform (40 Mal d1-, 3 Mal d3-, and 2 Mal d4-peptides; Fig. 1b). Forty-eight peptides could only be assigned to two Mal d isoforms (Supplementary Fig. S3a). Another 17 peptides could be assigned to a group of three isoforms and 20 to a group of four isoforms. Approximately 47% of all Mal d peptide sequences were contained in five or more allergens. The length of all identified Mal d peptides ranged from 7 to 37 amino acids (Fig. 1c). Peptides of 11, 14 and 15 amino acids were the most frequent. Sequence fragments from all sections of the Mal d1 primary protein structure were found; although for some positions within the amino acid sequence an increased number of peptides were detected such as positions 19–33, 41–55, 41–56 and 105–115 (Fig. 1d). The detected peptides of Mal d2 were distributed over the amino acid sequence with no peptides found in the region between 1 and 26 and 89–162 (Supplementary Fig. S3b). The Mal d3 peptides comprised amino acids 43–115, with positions 43–56, 57–63, and 77–96 in particular forming numerous peptides (Supplementary Fig. S3c). The Mal d4 sequence was nearly fully covered by the obtained peptides, which were frequently located in the regions 72–84, 96–121 and 122–131 (Supplementary Fig. S3d).

3.3. Hierarchical cluster analysis

Hierarchical cluster analysis was performed to visually differentiate genotypes based on their Mal d peptides composition using the detected intensities of all Mal d peptides (Supplementary Fig. S4). Similar genotypes gather within relatively small distances, whereas dissimilar genotypes are separated by relatively large gaps. For the apple samples derived from the years 2018 and 2019 the two biological replicates of each genotype clustered nicely together with a few exceptions (Supplementary Fig. S4b). In contrast, many replicates of the samples of 2017 did not always cluster together (Supplementary Fig. S4a). In the dendrogram well-tolerated genotypes (group I; Supplementary Fig. S2) are highlighted in blue, moderately tolerated genotypes (group II) are not marked, and allergenic genotypes (group III) are highlighted in red. For the genotypes harvested in 2017 there are in total two main clusters with corresponding subclusters (Supplementary Fig. S4a). Cluster 1 includes the allergenic genotypes p29 (p29-1, p29-2), p167-2 and p36-1, indicating a similar Mal d1-4 peptide composition for these genotypes. Subcluster 2.1 contains the poorly tolerated genotypes p211 (p211-1, p211-2), p167-1 and p36-2. This cluster does not only contain those poorly tolerated genotypes but also some of the well-tolerated genotypes p185 (p185-1, p185-2) and p10 (p10-1, p10-2). This demonstrates that even small variations in peptide composition or changes in abundance of some peptides, which can also be assigned to the allergens Mal d2-4, can drastically influence the overall tolerability of some apples.

The genotypes harvested in 2018 and 2019 cluster in two main clusters (Supplementary Fig. S4b). Few well-tolerated genotypes along with one moderately tolerated genotype can be found in cluster 1. A larger number of well-tolerated and all poorly tolerated genotypes are in cluster 2 but group in subclusters. Subcluster 2.1 contains the allergic symptoms triggering genotypes p78_2018 (p78-1, p78-2) and p211_2019 (p211-1, p211-2). Cluster 2.2 includes all the other poorly tolerated genotypes, of which GD (GD-1, GD-2) clusters together with p48 (p48-1, p48-2), and p17_2018 (p17-1, p17-2) clusters together with p36_2018 (p36-1, p36-2), indicating a similar peptide pattern, respectively. Some genotypes show similar proteomic profiles over two harvest years as it is the case for p211 (2018, 2019) located in subcluster 2.1 and SAN (2018, 2019) located in subcluster 2.2. Although the hypoallergenic variant SAN showed similar Mal d1-4 peptide profiles in the two harvest years, the mean symptom value obtained by oral provocation tests was higher in 2018 than in 2019. The genotype p92 of 2018 and 2019 (subcluster 2.2), as well as p78 of 2018 (subcluster 2.1) and 2019 (cluster 1) seem to contain contrasting Mal d1 compositions in the two years, as the samples of the different years are separated by large distances. The Mal d1 quality and quantity appears to be strongly affect in both genotypes by the harvest year. Although it is known that the allergic potential of different apple genotypes depends on the amount of Mal d proteins, these results show that there must be additional factors influencing allergenicity, because the well and poorly tolerated apples

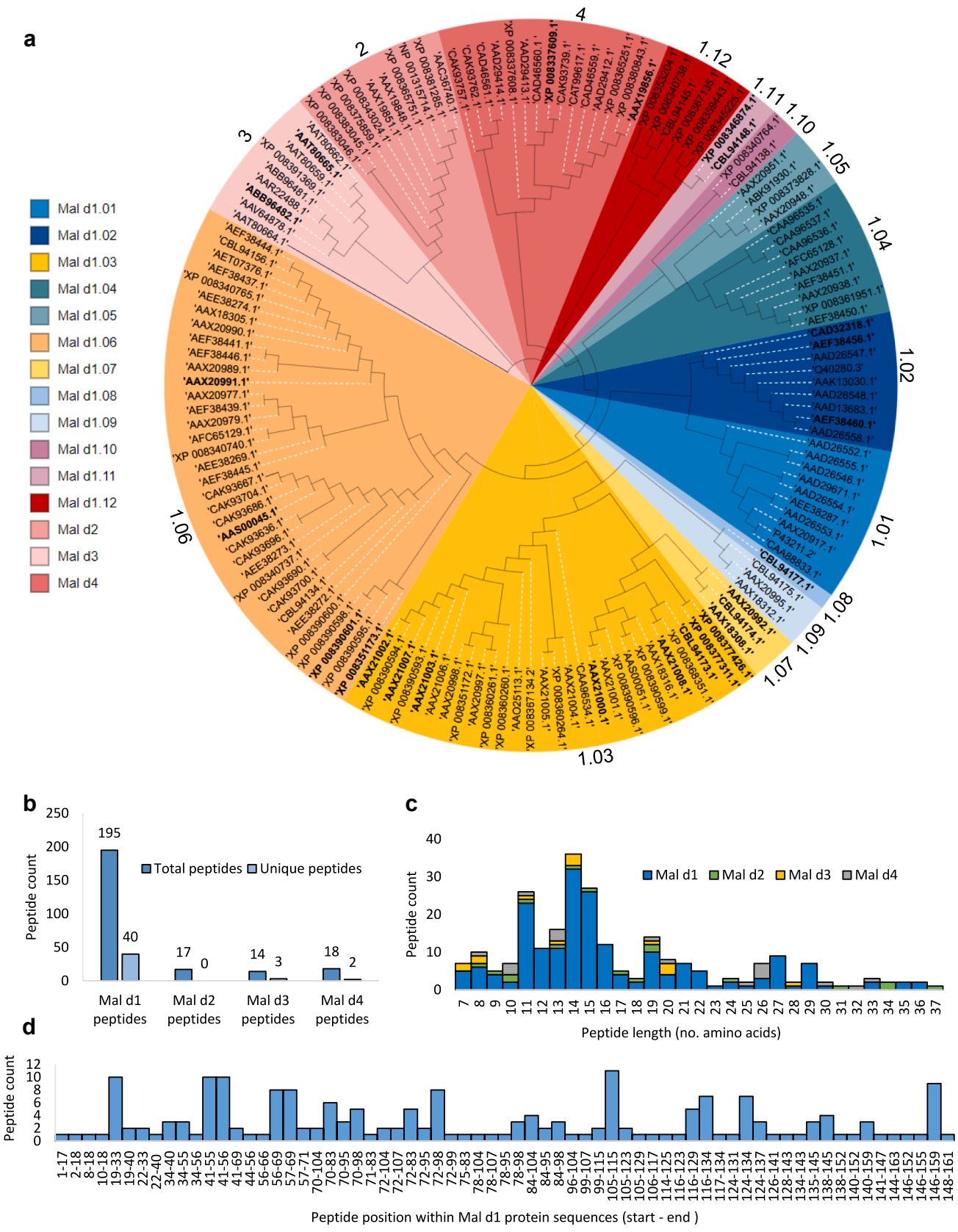

Fig. 1. NCBI-derived Mal d proteins and their isoforms. Phylogenetic analysis of Mal d proteins (refer to Material and Methods). Mal d group numbers are displayed in the outermost circle. Isoforms identified by at least one isoform-specific peptide are indicated in bold. Isoforms detected by unspecific peptides are highlighted with white dashed lines (**a**). Frequency of specific peptides for the corresponding allergens (**b**). Peptide length distribution (**c**). Peptide position within Mal d1 protein sequences (**d**).

Table 1
Allergens identified by isoform-specific peptides in 52 apple samples (Malus × domestica Borkh). Each isoform is presented in a different line. Related protein groups, the percentage of amino acid identity for Mal d1 isoforms to Bet v 1a and detected specific peptide sequences are presented.

Protein name/ NCBI accession number	Protein group	Pairwise identity to Bet v1 (%)	Matched specific peptides (Sequence position)
CAD32318.1	Mal d1.02	57	GVYTYENEYTSEIPPPR (2–18)
AEF38456.1	Mal d1.02	56	LFMAFVLDADNLIPK (19–33)
AEF38460.1	Mal d1.02	56	NITFGEGSQYGYVK (56–69)
XP_008377426.1	Mal d1.03	59	HKIDGVDKDNFVYK (70–83)
			LVAYGSGSVIK (105–115)
AAX21000.1	Mal d1.03	56	LMASGSGCVIK (105–115)
			HRIDGVDKDNLVYK (70–83)
			IDGVDKDNLVYK (72–83)
AAX21003.1	Mal d1.03	60	LVAAGSGSVIK (105–115)
AAX21002.1	Mal d1.03	60	LVAASSGSVIK (105–115)
AAX21008.1	Mal d1.03	59	LVASDNGSIIK (105–115)
			STSHYYTKGDVEIKEEHVK (116–134)
CBL94173.1	Mal d1.03	60	LVASDSGSIIK (105–115)
AAX21007.1	Mal d1.03	60	LVASSNGSVIK (105–115)
XP_008377311.1	Mal d1.03	57	YSVIEGDAITETIEK (84–98)
			YSVIEGDAITETIEKISYETK (84–104)
AAX20991.1	Mal d1.06	53	LYNALVLDADNLILK (19–33)
XP_008351173.1	Mal d1.06	56	HKVDGIDKDNFVYK (70–83)
			VDGIDKDNFVYK (72–83)
XP_008390601.1	Mal d1.06	56	STSHYHTKEDVEIK (116–129)
			STSHYHTKEDVEIKEEHVK (116–134)
			EDVEIKEEHVK (124–134)
AAS00045.1	Mal d1.06	55	LYYALVLDADNLLPK (19–33)
			VGKDKAHDLFK (135–145)
			DKAHDLFK (138–145)
AAX18308.1	Mal d1.07	60	HRIDGVDKENFVYQYSVIEGDAISETIEK (70–98)
CBL94174.1	Mal d1.07	60	HRIDGVDKDNFVYQYSVIEGDAISETIEK (70–98)
AAX20992.1	Mal d1.07	59	HKIDGVDKDNFVYQYSVIEGDAISETIEK (70–98)
CBL94177.1	Mal d1.08	58	LFNATALDGDELIAK (19–33)
			SIEILEGDGGVGTVQK (41–58)
			IIFGEGSTNGYVK (57–69)
			RIDVIDKDNFVYK (71–83)
			IDVIDKDNFVYK (72–83)
			ISYETTLVASGSGSIIK (99–115)
			GDVEINEEHLK (124–134)
CBL94148.1	Mal d1.11	33	MFNALILDAHNICPK (19–33)
			IDALDKEALSCTYTFIESDATDHLLDKLEYITYDVK (72–107)
			EALSCTYTFIESDATDHLLDKLEYITYDVK (78–107)
XP_008346874.1	Mal d1.11	34	MFNALILDSHNLCPK (19–33)
			IDALDKEALSCSYTFIESDASDHLMDKLEYITYDVK (72–107)
ABB96482.1	Mal d3	–	NGGAVPPACCNGIR (43–56)
			TINSLAR (57–63)
AAT80665.1	Mal d3	–	SLAGSVSGVNPGNVESLPGK (77–96)
AAX19856.1	Mal d4	–	KTGQDLVFGIYEEPLTPGQCNMIVER (96–121)
XP_008337609.1	Mal d4	–	LGDYLVEQGL (122–131)

do not clearly cluster according to their Mal d peptides and their abundances. Thus, different isoforms could cause different tolerances.

3.4. Allergen composition of apple genotypes

To determine the contribution of individual Mal d isoforms to allergenicity, the composition of Mal d isoforms in individual genotypes was examined in more detail. In the analyzed apple genotypes, 1 to 15 Mal d1 isoforms were expressed considering only those that were identified by specific peptides (Supplementary Table S2; Table 1). The genotype p193-2018 showed only non-specific Mal d1-derived peptides, which could not be assigned to a specific isoform. Mal d 1 isoforms share a pairwise identity from 93 to 99 % depending on the group of Mal d1 proteins (Supplementary Table S1). Consequently, in many cases it was not possible to distinguish between Mal d1 proteins based on peptides sequences, because the high sequence similarity resulted in a lack of isoform-specific peptides. Nevertheless, it is likely that multiple isoforms are expressed. For the group of Mal d1.01 proteins only unspecific peptides were detected (Fig. 1a). No peptide mapped to the proteins of subfamily Mal d1.12. Most Mal d1 proteins could be assigned to the Mal d1.03 protein group (Fig. 1a). The specific peptides derived from these proteins were predominantly located between amino acids 105–115 and differed in single amino acid exchanges (Table 1). The Mal d1 isoforms XP_008390601.1 (Mal d1.06, Fig. 1a) and AAX21002.1 (Mal d1.03, Fig. 1a), each identified by specific peptides (Table 1), were present in nearly all genotypes, indicating that they are two of the prevalent isoforms (Supplementary Table S2).

3.5. Mal d peptides differentially expressed between well and poorly tolerated apples

To identify potential allergenic peptides, we performed various T-tests to find peptides that are differentially expressed in a variety of genotypes, which are differently tolerated by allergic sufferers. T-test analysis was performed between well-tolerated genotypes (group I) and allergenic genotypes (group III), separated by harvest years. Fold changes were used for ranking differentially expressed peptides. Several peptides (p-value < 0.1, fold change ≥ 0.5) were more abundant in group III genotypes compared with group I genotypes (Fig. 2). The peptides DFDEPGSLAPTGLHLGGTK (amino acid (aa) 53–71), GDVEIKEEQVK (aa 124–134) and HRIDGLDKDNFVYK (aa 70–83) were 1.5, 2.8 and 9.5 times more abundant in group III apples than in group I apples of the harvest year 2017. These non-specific peptides can be assigned to small groups of protein isoforms (Fig. 2). Further peptides

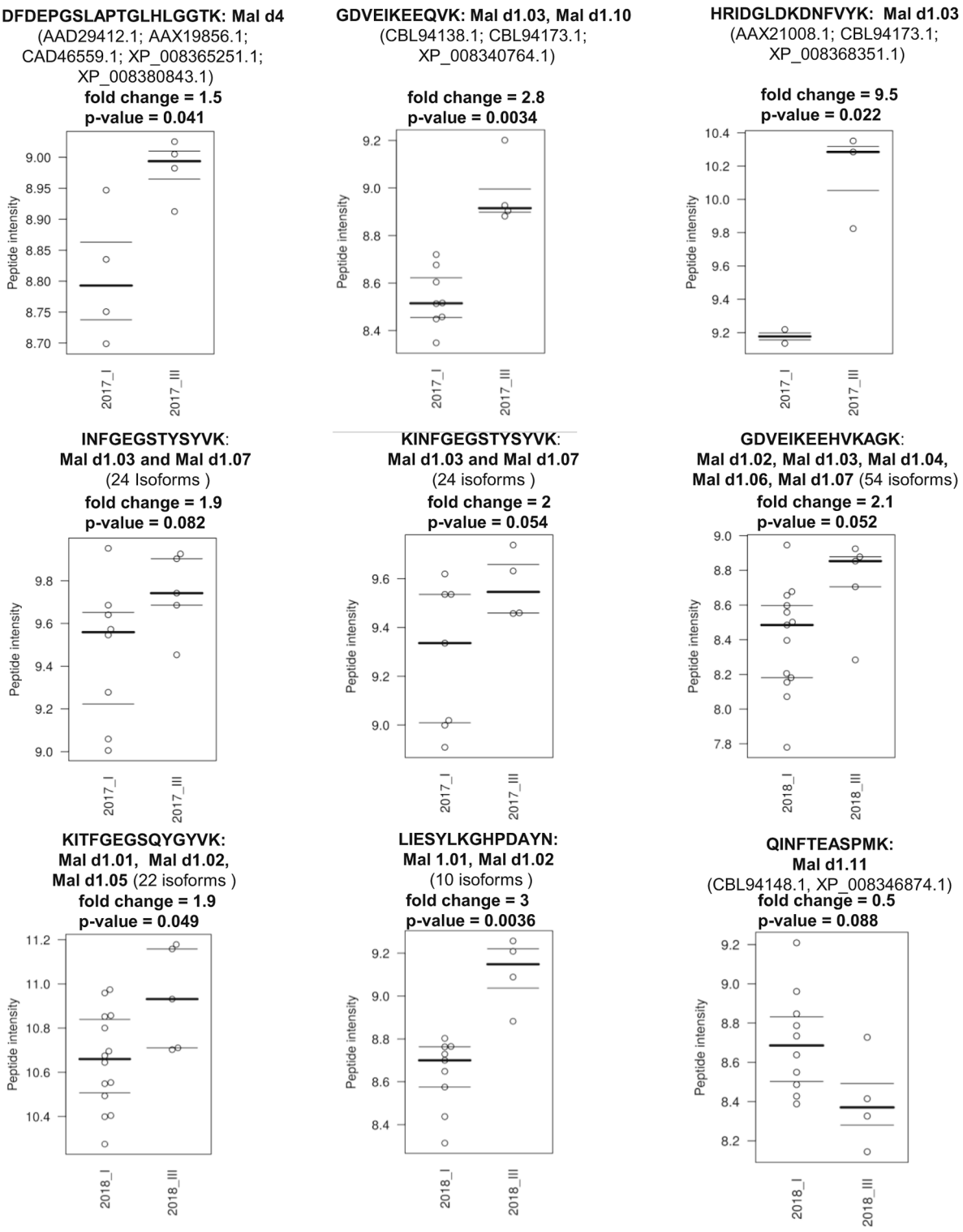

Fig. 2. Beeswarm Boxplot presentation of performed T-Test analysis between well-tolerated (group I) and allergenic genotypes (group III) of the harvest years 2017, 2018.

INFGEGSTYSYVK (aa 57–69), KINFGEGSTYSYVK (aa 56–69)) were overabundant in group III genotypes with fold changes of 1.9 and 2, respectively. Group III genotypes of 2018 also showed high levels of the three peptides GDVEIKEEHVKAGK (aa 124–137), KITFGEGSQYGYVK (aa 56–69) and LIESYLKGHPDAYN (aa 146–159) with fold changes of 2.1, 1.9 and 3, respectively compared to group I. The latter peptide can be assigned to a group of 10 Mal d1.01 and Mal d1.02 isoforms. The peptide QINFTEASPMK (aa 56–66) (fold change = 0.5) was slightly more expressed in Group I of 2018 compared to group III and can be assigned to the group of Mal d1.11 proteins.

The identification of differentially expressed Mal d protein isoforms requires detection of isoform-specific peptides. Therefore, we next focused on differentially expressed isoform-specific peptides. To account for the genetic diversity of apple genotypes, we compared the peptide

abundance of group I (well-tolerated genotypes) with individual genotypes in group III (allergenic genotypes) separately to identify peptides highly abundant in allergenic genotypes (Table 2). The Mal d1 isoform AAX21008.1 was reproducibly more abundant (fold change > 2.5) in four allergenic genotypes of group III (p36, p167, p29, p48) compared to group I (Table 2). This isoform was identified by two specific peptides LVASDNGSIIK (aa105-115) and STSHYYTKGDVEIKEEHVK (116–134) of which at least one of them was overabundant in each of the above-mentioned allergenic genotypes. The allergenic genotypes p29 (2017), p78 (2018) and p36 (2018) showed significantly higher levels of the peptides HKVDGIDKDNFVYK (aa 70–83) and VDGIDKDNFVYK (aa 72–83) that can be assigned to the Mal d1.06 isoform XP_008351173.1 (Table 2). The Mal d1.03 isoform AAX21002.1 identified by the detected specific peptide LVAASSGSVIK (105–115) was overabundant in the genotypes p29 (2017), p78 (2018) and p17 (2018) with fold changes exceeding 15. The specific peptide LYYALVLDADNLLPK (aa 19–33) linked to the isoform AAS00045.1 was more abundant in the cultivar GD in comparison to the group of well-tolerated apple genotypes (Table 2). Further allergenic genotypes p29 (2017) and p78 (2018) showed high levels of the specific peptides associated to AAS00045.1 with fold changes above 2.5 (Table 2). In contrast, some well-tolerated genotypes (e.g. p198 (2017), p10 (2017)) also showed high levels of the specific peptides for AAS00045.1 (Supplementary Table S3).

Next, the allergenic genotype (group III) were compared with individual well-tolerated genotypes (group I) to identify peptides that are reproducibly more abundant in the group I genotypes. The isoform CBL94148.1 was identified by three specific peptides MFNALILDAHNICPK (aa 19–33), IDALDKEALSCTYTFIESDATDHLLDKLEYITYDVK (aa 72–107) and EALSCTYTFIESDATDHLLDKLEYITYDVK (aa 78–107) (Table 1). Some well-tolerated genotypes harvested in 2018 showed significant high levels of the peptide MFNALILDAHNICPK (aa 19–33), linked to that isoform (Supplementary Table S4). Additionally, the isoform XP_008346874.1 identified by two specific peptides is highly expressed in three well-tolerated genotypes (p92-2018, p204-2018, p141-2018) (Supplementary Table S4). Both isoforms belong to the group of Mal d1.11 proteins (Fig. 1a). In contrast, in 2017, specific peptides for CBL94148.1 and XP_008346874.1 were not overabundant in well-tolerated genotypes (group I) (Supplementary Table S3). However, some of the well-tolerated apples showed high levels of non-specific peptides for the Mal d1.11 group to which both isoforms (CBL94148.1, XP_008346874.1) were assigned, indicating that they originate from different isoforms.

In addition to the Mal d1 isoforms, some Mal d2-4 proteins showed differential expression with fold changes exceeding 3.5 (Table 2, Supplementary Table 3, 4). In most cases, distinct protein assignment was not possible for Mal d2, Mal d3 and Mal d4 due to the lack of specific peptides. However, from the peptide analysis, it is evident that several Mal d2-4 proteins must be present. For Mal d2, no isoform-specific peptide was identified for the two proteins AAC36740.1 and NP_001315714.1 (Table 2, Supplementary Table S3, 4). These peptides were highly expressed in various genotypes such as p29 (2017), p78 (2017) (Table 2) and p13 (2017) (Supplementary Table S3). The Mal d3 proteins AAT80665.1 and ABB96482.1, each identified by specific peptides, were expressed in several genotypes (Table 2, Supplementary Table S3, 4). A specific peptide LGDYLVEQGL (aa 122–131) that is indicative for Mal d4 was significantly more expressed in the genotypes p15 (2018), p10 (2018), p185 (2018) and p36 (2018) (Supplementary Table S4, Table 2). Moreover, many of the well-tolerated genotypes in 2018 were also highly abundant in the non-specific peptide KSTMALLIGIYDEPMTPGQCNMVVER that can be associated to a group of two further Mal d4 proteins (Supplementary Table S4).

3.6. Correlation analysis identified peptides well correlated with allergenicity

Finally, we investigated the relationships between the peptide patterns and peptide intensity in each genotype and their Mal d1 content analyzed by ELISA (Fig. 3a, Supplementary Fig. S6a), as well as their correlations with the mean symptom score according to the human study with oral provocation tests on the other hand (Fig. 3b, Supplementary Fig. S6b). This correlation approach was employed to estimate the relative importance of each detected peptide and their corresponding isoforms. Correlation analyses between peptide intensity and Mal d1 content revealed several significant negative correlations. For the Mal d1 isoform CBL94177.1 two specific peptides were found and each of them showed a strong negative correlation with the allergen content with a correlation coefficient of −0.66 and −0.85 (Fig. 3a). In addition, the peptides found for CBL94177.1 showed a strong negative correlation with the mean symptom value (R = -0.63, −0.69, −0.78; Fig. 3b). In the case of the Mal d2 peptides, three peptides that can be assigned to the group of the two isoforms AAC36740.1 and NP_001315714.1, correlated significantly negatively (R = −0.65, −0.73, −0.82) with the Mal d1 content (Fig. 3a). One of those peptides (VCPAPLQVK) correlated also significantly negatively with the mean symptom value (-0.4; Supplementary Fig. S6b). For three peptides that can be assigned to small groups of Mal d3 isoforms, e.g. ABB96481.1, a significant negative correlation with Mal d1 content is shown (Fig. 3a). The peptide ISTSTNCATVK, specific for ABB96481.1 correlates negatively with the symptom score (Supplementary Fig. S6b). The negative correlation with the Mal d1 concentration for the two isoforms AAT80662.1 and AAT80659.1 was confirmed by two peptides (R= - 0.55, - 0.63; Supplementary Fig. S6a). The weakest significant correlation with the Mal d1 content was found for a group of Mal d4 proteins (R = −0.49; Supplementary Fig. S6a). No significant strong positive correlation (R > 0.5) was calculated between the intensities of Mal d1-derived peptides and Mal d1 content or mean symptom score.

4. Discussion

The apple genome contains the genetic information for various Mal d1 proteins (paralogs) due to multiple duplications of the coding genes, which also occur in allelic variants in the diploid organism. Thus, individual apple genotypes possess a set of Mal d1 isoforms, which also occur at different frequencies in different apple varieties. This biodiversity can be exploited for selecting genotypes with reduced allergenicity. In this study, we analyzed 52 apple samples of various apple genotypes, previously evaluated in a human provocation test (Romer et al., 2020), for their proteomic profile regarding the abundance of different apple allergens and their isoforms using proteomics and bioinformatic approaches.

4.1. Hypo- and hyperallergenic variants of Mal d1

In recent years, differences in the IgE binding ability of the Mal d1 homolog Bet v1 and its isoforms have already been demonstrated by immunoblot experiments. For Bet v1, nine isoforms could be grouped into hypoallergenic and allergenic variants according to their IgE reactivity (Ferreira et al., 1996; Swoboda et al., 1995). Many Mal d1 isoforms have been described in previous studies at the gene level but their presence at the protein level have not been confirmed in most cases (Puehringer, 2003; Gao et al., 2005). So far, a few Mal d1 genes and their paralogous and allelic variants are known that have been associated with low or high allergenic potential. By site-directed mutagenesis mutant Mal d1 proteins have been assessed for hypoallergenicity (Son et al., 1999; Bolhaar et al., 2005b). Several hypoallergenic and hyperallergenic variants of Mal d1 have been postulated (Supplementary Table S5) (Vegro et al., 2016; Pagliarani et al., 2012; Gao et al., 2008). The critical factors that cause some isoforms to be less IgE-reactive and others to be more IgE-reactive have not been fully investigated.

Table 2
Peptides overabundant in allergenic apple genotypes of all three harvest years. Overabundant peptides with a statistically significant p-value (<0.05), resulting from a T-Test analysis comparing the peptide abundance of each group III genotype to that of group I, are shown. Only positive fold changes indicating an increase in abundancy of the presented peptides for each allergenic genotype (group III) are shown. Specific peptides are shown in bold font.

		Harvest year	2017					2018					2019
		Genotypes	p160	p186	p36	p167	p29	p48	p78	GD	p17	p36	p186
		Mal d1 [µg/g FW]	-	-	-	-	-	-	-	-	8.33 ± 1.10	-	0.67 ± 0.31
Peptide sequence (position within the protein sequence)	Protein name/NCBI accession number	Protein group											
HKVDGIDKDNFVYK (70–83)	XP_008351173.1	Mal d1.06					44.67		5.37			12.02	
VDGIDKDNFVYK (72–83)							2.51					4.47	
LVASSNGSVIK (105–115)	AAX21007.1	Mal d1.03										4.17	
LVASDNGSIIK (105–115)	AAX21008.1	Mal d1.03			3.98	4.79	29.51	3.98					
STSHYYTKGDVEIKEEHVK (116–134)						2.57							
LYYALVLDADNLLPK (19–33)	AAS00045.1	Mal d1.06							2.82	5.62			
DKAHDLFK (138–145)							3.02						
LVAASSGSVIK (105–115)	AAX21002.1	Mal d1.03					58.88		15.49		18.20		
LVAAGSGSVIK (105–115)	AAX21003.1	Mal d1.03						4.27					
GDFEIKEK (124–131)	AEF38445.1	Mal d1.06					3.31						
	AFC65129.1	Mal d1.06											
TVEILEGDGSVGTIK (41–55)	AAS00045.1	Mal d1.06		4.17			5.50		3.72	3.63		3.09	
TVEILEGDGSVGTIKK (41–56)	AAX20977.1	Mal d1.03		12.59			8.91		5.31	4.37			
VTFGEGSQLGFVK (57–69)	CBL94138.1	Mal d1.10			2.57								
LVASPDGGSIVK (106–117)	XP_008340764.1	Mal d1.10				2.88					3.55		
SDVEIKEEHVK (124–134)	AAX20997.1	Mal d1.03						2.95					
	CAA96534.1	Mal d1.03											
LIESYLKDHPDAYN (146–159)	AAD26552.1	Mal d1.01		16.98	11.22	11.48						2.82	
	AAD26553.1	Mal d1.01											
	AAD26554.1	Mal d1.01											
	AAD29671.1	Mal d1.01											
	AAX20917.1	Mal d1.01											
	AEE38287.1	Mal d1.01											
	P43211.2	Mal d1.01											
	CAD32318.1	Mal d1.02											
QAEILEGNGGPGTIK (41–55)	AAD26552.1	Mal d1.01					12.30					3.09	
QAEILEGNGGPGTIKK (41–56)	AAD29671.1	Mal d1.01		5.50			7.76					3.31	
	AAD26554.1	Mal d1.01											
	AAD26553.1	Mal d1.01											
	AEE38287.1	Mal d1.01											
	AAD26546.1	Mal d1.01											
	AAD26555.1	Mal d1.01											
	AAX20917.1	Mal d1.01											
	CAA88833.1	Mal d1.01											
	P43211.2	Mal d1.01											
VCPAPLQVK (162–170)	AAC36740.1	Mal d2					11.30		3.31				
SACLAFGDSKYCCTPPNNTPETCPPTEYSEIFEK (181–214)	NP_001315714.1								29.51				2.63
YCCTPPNNTPETCPPTEYSEIFEK (191–214)									6.92				
ITFTNNCPNTVWPGTLTGDQKPQLSLTGFELASKASR (27–63)													2.51
SLAGSVSGVNPGNVESLPGK (77–96)	AAT80665.1	Mal d3									5.75		
LGDYLVEQGL (122–131)	XP_008337609.1	Mal d4										3.24	
KTGQALVFGIYEEPLTPGQCNMIVER (96–121)	XP_008365251.1	Mal d4									4.90		
	AAD29412.1	Mal d4											

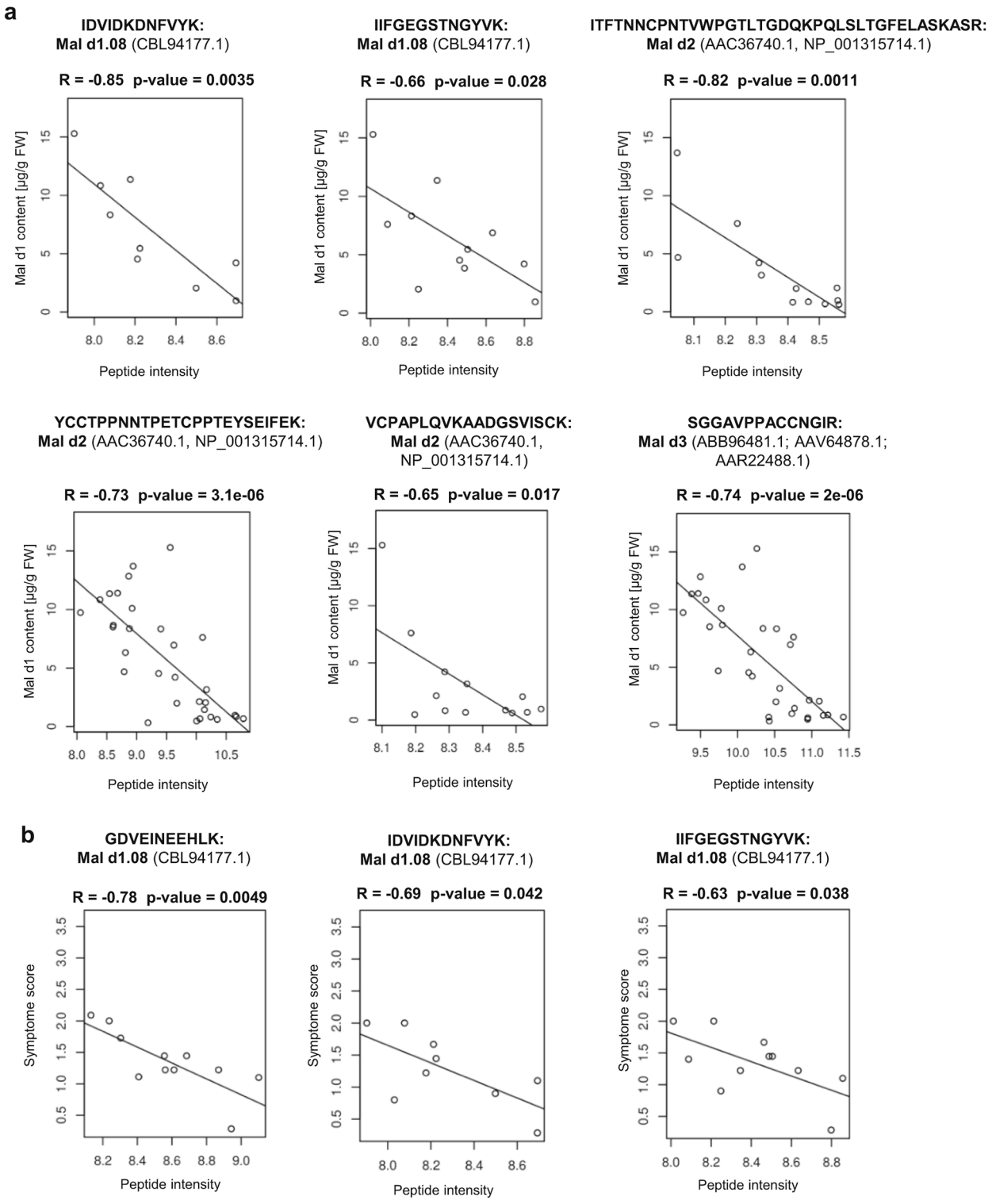

Fig. 3. Correlation analysis showing the relationship between peptide intensity and the Mal d1 content (**a**) and the relationship between peptide intensity and the mean symptom value according to the human study with oral provocation tests (**b**).

4.2. Mal d1 isoforms in apple

Our proteomic analysis revealed significant differences in the peptide profiles of the analyzed apple genotypes that included well-tolerated and allergic symptoms triggering genotypes. Mal d1 comprises at least 113 variants, whereas in our study 21 isoforms were clearly confirmed by the detection of at least one isoform-specific peptide. Other isoforms are probably also expressed but could not be positively verified due to lack of specific peptides. This implies that apple eaters are usually exposed to a variety of different Mal d1 isoforms with varying IgE-reactivity and it is supposed that not all expressed isoforms are of clinical relevance. We demonstrated that at least 1 to 15 Mal d1 isoforms are expressed in each genotype. Similarly, for birch species the expression of four to six isoforms per specie has been

reported and the presence of further 15 isoforms has been supposed (Schenk et al., 2009).

Preliminary studies revealed that Mal d1a (NCBI accession number: P43211.2; Mal d1.01; Fig. 1a) and Mal d1b (NCBI accession number: Q40280.3; Mal d1.02) isoforms are the most abundant isoforms present in apples at the transcriptome level (Son et al., 1999). Since these isoforms do not provide specific peptides by tryptic digestion, they could not be unequivocally detected in our study, but 23 and 29 peptides could be identified for P43211.2 and Q40280.3, respectively that match the protein sequence of these Mal d1 isoforms (Fig. 4b, Supplementary Fig. S7). Members of the protein families Mal d1.01 and Mal d1.02 have been associated with hyperallergenic properties because their genes were transcribed less frequently in hypoallergenic varieties (Paris et al., 2017). While Mal d1.01 and Mal d1.02 are among the most abundantly expressed isoforms, members of the group Mal d1.12 are the least expressed isoforms, consistent with our results, as no peptide for this protein group was detected (Fig. 1a) ((Paris et al., 2017).

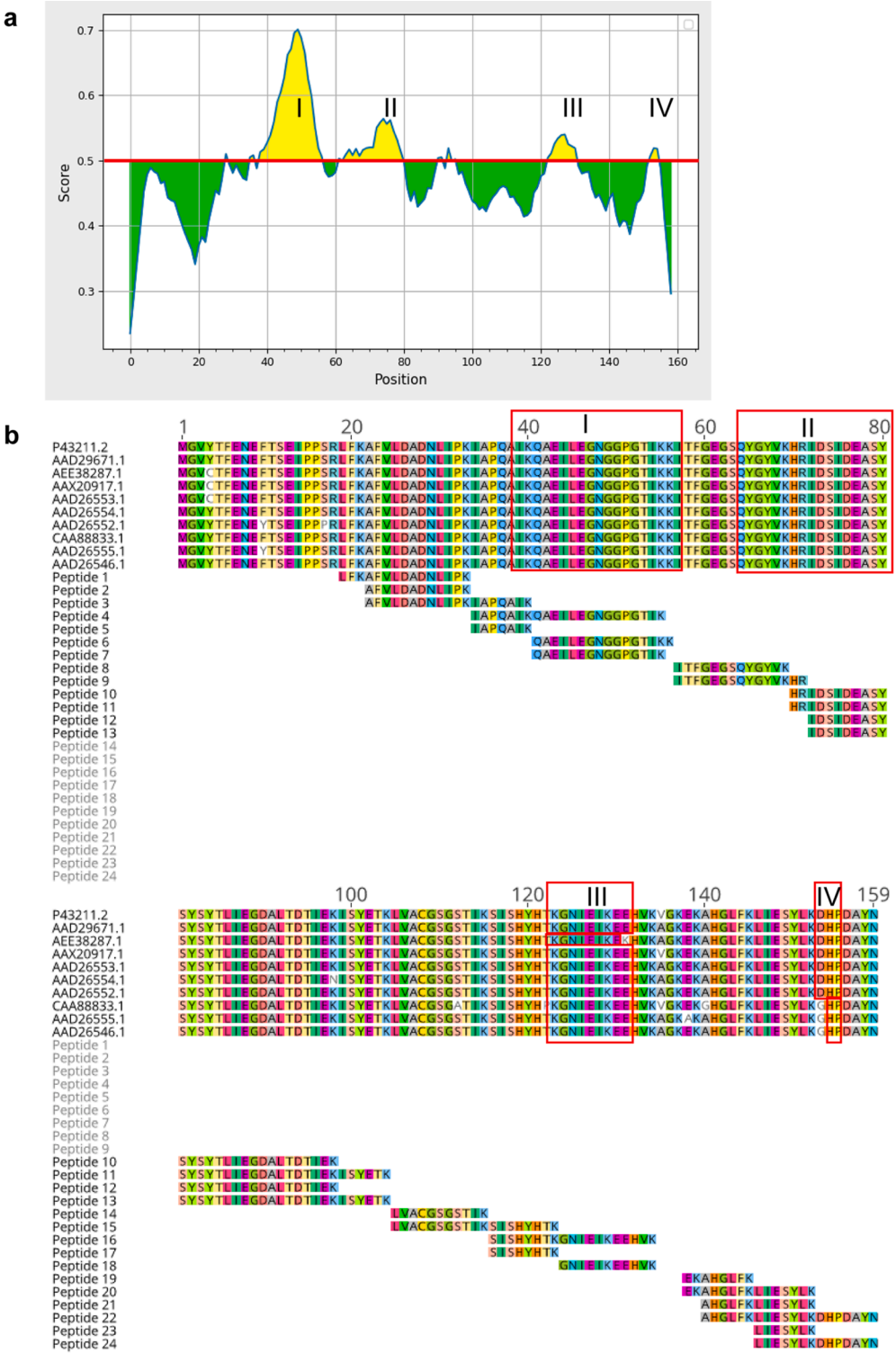

Fig. 4. Predicted epitopes of the isoform AAD29671.1 of the group of Mal d1.01 proteins. Regions colored in yellow (IIV) in the graph are predicted epitopes using the iedb.org online tool (http://tools.iedb.org/bcell/). The default threshold of 0.5 is highlighted with a red line (**a**). Amino acid sequence alignment of Mal d1.01 proteins and identified peptides by LC-MS (Image was generated with Geneious Prime (v. 2021.1.1) (https://www.geneious.com)). Predicted epitopes (IIV), referring to (**a**) are marked by red boxes (**b**). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Structural features of Mal d1 proteins functioning as epitope

Mal d1 and other PR-10 proteins share a common structural motif including the highly conserved region with the sequence G-X-G-G-X-G-X (aa 46–52), referred to as the phosphate loop (Fig. 4b). For Bet v1, it was shown that the Fab fragment of a monoclonal murine IgG1 antibody BV16 binds to Bet v1 in the P-loop region (Mirza et al., 2000). The residue Glu-45, which is involved in hydrogen binding to the fragment, seems to play a key role in the binding mechanism as a substitution of Glu-45 by serine led in a 50% reduction of the binding capacity, indicating that the P-loop motif is one important epitope (Spangfort et al., 2003). The importance of this motif as a B-cell epitope has also been described for the Bet v1 homologous protein Pru av1, the major cherry allergen (Neudecker et al., 2003). In our study several peptides were identified, such as TVEILEGDGSVGTIK (aa 40–55), TVEILEGDGSVGTIKK (aa 40–56) and GIEILEGNGGVGTIK (aa 40–55) that contained the conserved glycine-rich motif including Glu-45. Similar peptides were found in a preliminary proteomic study, and were attributed an important role in binding of human IgE (Romer et al., 2020). However, two isoforms of Mal d1 do not bind to BV16, although they both contain the conserved Glu-45 residue (Holm et al., 2001). Therefore, other amino acid residues are also relevant for IgE binding. A site-directed mutagenesis study revealed that amino acid residues Thr10, Thr57, Ser111 and Thr112 are crucial for IgE reactivity (Supplementary Table S6) (Son et al., 1999; Ma et al., 2006). Analysis of these amino acids in isoforms identified by specific peptides in this study showed that there are three isoforms (AEF38456.1, AEF38460.1 and CAD32318.1) that contain three of the crucial amino acid residues namely Thr10, Thr57, and Ser111 (Supplementary Table S6). It is striking that none of the detected proteins contained all four residues in their primary structure. The residues Ser111 and Thr10 characterize the Mal d1 isoforms AAX21002.1 and AAX21008.1, which are prevalent in allergenic genotypes (Table 2). Further isoforms, such as CBL94138.1, CBL94173.1, XP_008340764.1 and XP_008368351.1, of which at least one is expressed and has been shown to be more abundant in allergenic genotypes, carry these crucial amino acid residues too (Fig. 2).

In contrast, the Mal d1 isoforms CBL94148.1 and XP_008346874.1, which are prevalent in well-tolerated genotypes (Fig. 2, Supplementary Table S3, 4) do not contain any of the mentioned crucial amino acids and further show a low sequence similarity to Bet v1 with only 33 and 34 %, respectively. This makes the occurrence of common IgE-epitopes unlikely and may explain the good tolerance of genotypes, which express these isoforms (Supplementary Table S3, 4).

Oligomerization of allergens can reduce IgE binding and thus allergenicity (Pagliarani et al., 2012). A Ser113Cys change in a Bet v1 isoform has previously been identified as being important for the ability of Bet v1 to form aggregates, possibly through the formation of intermolecular disulphide bonds, and to create a type of protection against IgE binding (Zaborsky et al., 2010). Similar properties can also be attributed to the Mal d1.11 isoforms and the isoform AAX21000.1 (from the Mal d 1.03C02 allele), which have a Cys at a similar position (Pagliarani et al., 2012). Although AAX21000.1 shows a comparable high sequence similarity to Bet v1 with 56 %, it is noticeable that it does not contain any of the crucial amino acid residues and hence might be affiliated with a lower IgE binding capacity (Supplementary Table S6).

The protein CBL94177.1 appears to be a hypoallergenic isoform, as several negative correlations were found between the associated peptides of that isoform with the allergen content as well as with the mean symptom score (Fig. 3a, b). The occurrence of CBL94177.1 in the known, well-tolerated cultivar SAN confirms this hypothesis (Romer et al., 2020). Even if this protein can be associated with hypoallergenic properties; it is evident that isoforms that are hypoallergenic to some individuals may still be allergenic to others (Schenk et al., 2008).

In contrast, the allele Mal d1.06A01 was present alone or together with Mal d1.06A03 in hyperallergenic cultivars such as GD, whereas the allele Mal d1.06A02 was detected only in the well-tolerated variety SAN (Gao et al., 2008). As additional endorsement a significant positive correlation for the allele Mal d1.06A01 with immune-reactivity of patients sera was recently presented (Siekierzynska et al., 2021). The three variants of Mal d1.06A each differ in one single hydrophobic amino acid 13 V/I and 135 V/A (Gao et al., 2008). In addition, the isoform AAS00045.1, which belongs to the group of Mal d1.06 proteins, was highly expressed in the poorly tolerated cultivar GD (2018) (Table 2). This is in accordance with the findings of a recent proteomic study and demonstrates that the accumulation of some isoforms can be stable within the variety over different harvest years (Romer et al., 2020). Although this isoform does not originate from the Mal d1.06A01 allele, it has a valine residue at position 13 and 135, like the isoforms of Mal d1.06A01, that may be important for epitope conformation (Gao et al., 2008). In our study, several well-tolerated genotypes also accumulated this isoform (Supplementary Table S3, 4). Therefore, it can be assumed that some isoforms may have higher IgE-binding activity, but their role in triggering allergic responses also depends on the expression level in the genotype. Consequently, the abundancy of some isoforms will strongly affect the overall allergenicity of a genotype.

For Bet v1 two major T cell epitopes have been reported, which are located between amino acids 112–123 and 142–156 (Jahn-Schmid et al., 2005). Similarly, for peptide LFKLIESYLKDHPDAYN located in the region covering amino acids 142–158 in Mal d1, T cell activation was identified (Geroldinger-Simic et al., 2013). In our study we found two similar peptides AHGLFKLIESYLKDHPDAYN (aa 140–159) (Supplementary Table S2) and LIESYLKDHPDAYN (aa 146–159) (Table 2) that are located within the reported region and are specific for several proteins of the Mal d1.01 group. The peptide LIESYLKDHPDAYN was significantly highly abundant with fold changes > 10 in three allergenic genotypes of the harvest year 2017 (p211, p36, p167) (Table 2) and with a fold change of 2.82 in p36 in 2018. Another identified T cell epitope in Mal d1 is the sequence QAEILEGNGGPG located between aa 40 to 51 in the highly conserved glycine rich loop region (Geroldinger-Simic et al., 2013). We detected two similar peptides (Table 2), that differed only in the length to the reported one and can be assigned to the group of Mal d1.01 proteins. These short peptide sequences further seemed to have sequential B cell epitope properties as predicted by the antibody epitope prediction tool (Fig. 4a) (https://tools.iedb.org/bcell/) (Jespersen et al., 2017). Furthermore, a B cell epitope of Bet v1 covering the sequence NIEGNGGPGTIKK in the region 44–56 has been reported (Zhang et al., 2018) showing ~ 90% similarity to the Mal d1 peptides presented here (https://www.iedb.org, Organism: *Betula verrucosa*). Ultimately, it can be assumed that isoforms containing those highlighted peptides in the corresponding region can be associated with a higher allergic potential. Conversely, it must be taken into account that T-cell epitopes may vary from patient to patient and depend on the human leukocyte antigen. (Uehara et al., 2001).

4.4. Mal d2-4 proteins

In addition to Mal d1 other allergens have also been identified in this study, although the clinical relevance of Mal d2, Mal d3 and Mal d4 is not fully investigated yet. Mal d2 was detected by five isoform unspecific peptides and was assigned to a group of two thaumatin-like proteins (AAC36740.1, NP_001315714.1) (Table 2, Supplementary Table S3, 4). Of them, NP_001315714.1 was already confirmed in a preliminary study (Guarino et al., 2007) and another study showed that the impact of Mal d2 on the allergenicity of one cultivar is low, compared to other allergens (Fernández-Rivas et al., 2006). Refolding studies of Mal d2 led to the suggestion that IgE-binding epitopes are mostly hidden inside of the protein and making it difficult for IgE-antibodies to access them (Marzban et al., 2009). For Mal d3, a non-specific lipid transfer protein (nLTP), whose expression is triggered by various environmental conditions and increases at the end of maturity and during storage, we identified two proteins, which are AAT80665.1 and ABB96482.1 (Table 1) (Sancho et al., 2006). Furthermore, many non-specific Mal d3-

derived peptides have been detected, indicating a broad occurrence of Mal d3 proteins (Supplementary Table S2). It was reported that patients sensitized to Bet v1 showed a 3.5-fold lower risk of an IgE-mediated reaction to Mal d3 (Fernández-Rivas et al., 2006). Another allergen identified is the profilin Mal d4, a cytosolic protein, involved in intracellular transport processes and playing a key role in cell elongation and cell shape maintenance (Witke, 2004). In our study, we clearly identified the proteins XP_008337609.1 and AAX19856.1 (Table 1). Symptoms to Mal d4 tend to be mild with IgE-mediated OAS being the most prevalent one, as profilins are sensitive to heat and pepsin digestion (Zuidmeer et al., 2008). IgE binding ability for profilins was observed in only 10 to 30% of patients sensitized to Bet v1, suggesting that they play a minor role in birch Rosaceae-fruit syndrome (Breiteneder & Radauer, 2004). This is in accordance with our findings, as Mal d4 associated peptides did correlate positively with the mean symptom value of Bet v1- and Mal d1- sensitized patients (Supplementary Fig S6b). It must be noted that the presence of Mal d2, Mal d3 and Mal d4 cannot be directly linked to the symptom values measured in our human study, as all participating patients were sensitized to Bet v1 and allergic to apples, indicating that symptoms arose primarily from IgE-reactivity of Mal d1 proteins. However, their presence may have contributed to some extent to the allergic response and severity of the symptoms.

4.5. Limitations of the methods used

Proteomic analysis of allergenic protein isoforms in apples proved challenging. Unambiguous isoform assignment relies on the detection of isoform-specific peptides. This is not always given, as loss of specific peptides can occur due to various methodological obstacles such as sample preparation, peptide ionization and mass range limitations. As isoforms frequently share a high percentage of their amino acid sequences, the number of isoform-specific peptides is often limited to one or very few peptides per protein isoform. Further, when working with a protease like trypsin during sample preparation, achieving a complete protein sequence coverage is often not possible, as certain sequence areas might not have enough, or too many, trypsin cleavage sites and therefore do not deliver a peptide in a suitable size for mass spectrometry. Furthermore, identification of Mal d peptides is highly dependent on the quality of the apple reference proteome, particularly on the precision of the genome annotation on the isoform level. Moreover, it should be noted that the performed differential analysis is based on groups of genotypes with their corresponding symptom values provided by a human study. It needs to be highlighted that the mean symptom value was obtained by self-reported allergic symptoms of allergic patients. Additionally, the observed symptom scores for several genotypes often showed a high variance, including various outliers (Supplementary Fig. S1). This might have led to an increase of the mean symptom scores for some genotypes, despite being tolerated by a large proportion of patients, as it was exemplified for genotype p211 in the harvest years 2017 and 2019. Consequently, it is possible that the subdivision of genotypes based on their symptom values itself but also by using the mean symptom score, can be afflicted with errors and might have affected the results of the statistical analysis. Therefore, further studies are needed to assess the IgE-reactivity of peptide markers and proteins that have been identified in this study in terms of hypoallergenic and allergenic properties.

5. Conclusion

The proteomics approach used in this study allowed fast screening of Mal d1 isoforms and other apple allergens by quantifying the abundance of distinct protein isoforms (Levin, 2011; Bubis et al., 2017). We detected 21, 2, and 2 specific isoforms of Mal d1, Mal d3, and Mal d4 proteins, respectively in 52 apple samples of different genotypes. Our data identified several specific and unspecific peptides linked to various isoforms that might be possible markers to distinguish between well and poorly tolerated apple cultivars. The group of Mal d1.01 proteins may have a high impact on apple allergenicity, as Mal d1.01 peptides (aa 41–56) affiliated with T cell and predicted B cell properties were identified. In contrast, specific peptides of the isoform CBL94177.1 correlated negatively with the mean symptom value, indicating hypoallergenic properties of this isoform. In addition, high levels of the isoforms CBL94148.1 and XP_008346874.1 may also indicate lower allergenicity of apple cultivars. Ultimately, the identified peptides can be used for further assessment of hypoallergenicity using skin prick tests. It is evident, that this knowledge will also be important for the development of suitable recombinant proteins for immunotherapy of allergic patients or the breeding of hypoallergenic apple cultivars.

6. Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD027753.

Funding

This work was supported by the Federal Ministry of Food and Agriculture (Germany) (Project number: 2814IP018).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Katrin Plate for the preparation of the fruit material. We would like to acknowledge Miriam Abele for mass spectrometric support and Franziska Hackbarth for expert technical assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochms.2022.100111.

References

Beuning, L., Bowen, J., Persson, H., Barraclough, D., Bulley, S., & Macrae, E. (2004). Characterisation of Mal d 1-related genes in Malus. *Plant Molecular Biology, 55*, 369–388.

Bolhaar, S. T. H. P., van de Weg, W. E., van Ree, R., Gonzalez-Mancebo, E., Zuidmeer, L., Bruijnzeel-Koomen, C. A. F. M., … Gilissen, L. J. W. J. (2005a). In vivo assessment with prick-to-prick testing and double-blind, placebo-controlled food challenge of allergenicity of apple cultivars. *The Journal of Allergy and Clinical Immunology, 116*, 1080–1086.

Bolhaar, S. T. H. P., Zuidmeer, L., Ma, Y., Ferreira, F., Bruijnzeel-Koomen, C. A. F. M., Hoffmann-Sommergruber, K., … Knulst, A. C. (2005b). A mutant of the major apple allergen, Mal d 1, demonstrating hypo-allergenicity in the target organ by double-blind placebo-controlled food challenge. *Clinical Experimental Allergy, 35*, 1638–1644.

Breiteneder, H., & Radauer, C. (2004). A classification of plant food allergens. *The Journal of Allergy and Clinical Immunology, 113*, 821–830. quiz 831.

Bubis, J. A., Levitsky, L. I., Ivanov, M. V., Tarasova, I. A., & Gorshkov, M. V. (2017). Comparative evaluation of label-free quantification methods for shotgun proteomics. *Rapid Communications in Mass Spectrometry, 31*, 606–612.

Burney, P. G. J., Potts, J., Kummeling, I., Mills, E. N. C., Clausen, M., Dubakiene, R., … van Ree, R. (2014). The prevalence and distribution of food sensitization in European adults. *Allergy, 69*, 365–371.

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology, 26*, 1367–1372.

Fernández-Rivas, M., Bolhaar, S., González-Mancebo, E., Asero, R., van Leeuwen, A., Bohle, B., … van Ree, R. (2006). Apple allergy across Europe: How allergen sensitization profiles determine the clinical expression of allergies to plant foods. *The Journal of Allergy and Clinical Immunology, 118*, 481–488.

Ferreira, F., Hirtenlehner, K., Jilek, A., Godnik-Cvar, J., Breiteneder, H., Grimm, R., … 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. *The Journal of Experimental Medicine, 183*, 599–609.
Gao, Z., van de Weg, E. W., Matos, C. I., Arens, P., Bolhaar, S. T. H. P., Knulst, A. C., … Gilissen, L. J. W. J. (2008). Assessment of allelic diversity in intron-containing Mal d 1 genes and their association to apple allergenicity. *BMC Plant Biology, 8*, 116.
Gao, Z. S., van de Weg, W. E., Schaart, J. G., Schouten, H. J., Tran, D. H., Kodde, L. P., … Gilissen, L. J. W. J. (2005). Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (Malus domestica). *Theoretical and Applied Genetics, 111*, 171–183.
Geroldinger-Simic, M., Kinaciyan, T., Nagl, B., Baumgartner-Durchschlag, U., Huber, H., Ebner, C., … Bohle, B. (2013). Oral exposure to Mal d 1 affects the immune response in patients with birch pollen allergy. *The Journal of Allergy and Clinical Immunology, 131*, 94–102.
Guarino, C., Arena, S., de Simone, L., D'Ambrosio, C., Santoro, S., Rocco, M., … Marra, M. (2007). Proteomic analysis of the major soluble components in Annurca apple flesh. *Molecular Nutrition & Food Research, 51*, 255–262.
Holm, J., Bærentzen, G., Gajhede, M., Ipsen, H., Larsen, J. N., Løwenstein, H., … Spangfort, M. D. (2001). Molecular basis of allergic cross-reactivity between group 1 major allergens from birch and apple. *Journal of Chromatography B: Biomedical Sciences and Applications, 756*, 307–313.
Jahn-Schmid, B., Radakovics, A., Lüttkopf, D., Scheurer, S., Vieths, S., Ebner, C., & Bohle, B. (2005). Bet v 1142–156 is the dominant T-cell epitope of the major birch pollen allergen and important for cross-reactivity with Bet v 1-related food allergens. *Journal of Allergy and Clinical Immunology, 116*, 213–219.
Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research, 45*, W24–W29.
Lerma-García, M. J., Nicoletti, M., Simó-Alfonso, E. F., Righetti, P. G., & Fasoli, E. (2019). Proteomic fingerprinting of apple fruit, juice, and cider via combinatorial peptide ligand libraries and MS analysis. *Electrophoresis, 40*, 266–271.
Levin, Y. (2011). The role of statistical power analysis in quantitative proteomics. *Proteomics, 11*, 2565–2567.
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*, 53–62.
Marzban, G., Herndl, A., Pietrozotto, S., Banerjee, S., Obinger, C., Maghuly, F., … Laimer, M. (2009). Conformational changes of Mal d 2, a thaumatin-like apple allergen, induced by food processing. *Food Chemistry, 112*, 803–811.
Matricardi, P. M., Kleine-Tebbe, J., Hoffmann, H. J., Valenta, R., Hilger, C., Hofmaier, S., … Ollert, M. (2016). EAACI Molecular Allergology User's Guide. *Pediatric Allergy and Immunology: Official Publication of the European Society of Pediatric Allergy and Immunology, 27*(Suppl 23), 1–250.
Mirza, O., Henriksen, A., Ipsen, H., Larsen, J. N., Wissenbach, M., Spangfort, M. D., & Gajhede, M. (2000). Dominant epitopes and allergic cross-reactivity: complex formation between a Fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v 1. *Journal of Immunology (Baltimore, Md.: 1950), 165*, 331–338.
Neudecker, P., Lehmann, K., Nerkamp, J., Haase, T., Wangorsch, A., Fötisch, 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*, 97–107.
Pagliarani, G., Paris, R., Arens, P., Tartarini, S., Ricci, G., Smulders, M. M. J., & van de Weg, W. E. (2013). A qRT-PCR assay for the expression of all Mal d 1 isoallergen genes. *BMC Plant Biology, 13*, 51.
Pagliarani, G., Paris, R., Iorio, A. R., Tartarini, S., Del Duca, S., Arens, P., … van de Weg, E. (2012). Genomic organisation of the Mal d 1 gene cluster on linkage group 16 in apple. *Molecular Breeding New Strategies in Plant Improvement, 29*, 759–778.
Paris, R., Pagliarani, G., Savazzini, F., Aloisi, I., Iorio, R. A., Tartarini, S., … Del Duca, S. (2017). Comparative analysis of allergen genes and pro-inflammatory factors in pollen and fruit of apple varieties. *Plant Science: An International Journal of Experimental Plant Biology, 264*, 57–68.
Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., … Vizcaíno, J. A. (2019). The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Research, 47*, D442–D450.
Potocnakova, L., Bhide, M., & Pulzova, L. B. (2016). An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. *Journal of Immunology Research, 2016*, 6760830.
Puehringer, H. (2003). MdAP, a novel protein in apple, is associated with the major allergen Mal d 1. *Gene, 321*, 173–183.
Romer, E., Chebib, S., Bergmann, K.-C., Plate, K., Becker, S., Ludwig, C., … Schwab, W. (2020). Tiered approach for the identification of Mal d 1 reduced, well tolerated apple genotypes. *Scientific Reports, 10*, 9144.
Sánchez-Monge, R., Lombardero, M., García-Sellés, F. J., Barber, D., & Salcedo, G. (1999). Lipid-transfer proteins are relevant allergens in fruit allergy. *Journal of Allergy and Clinical Immunology, 103*, 514–519.
Sanchez-Trincado, J. L., Gomez-Perosanz, M., & Reche, P. A. (2017). Fundamentals and Methods for T- and B-Cell Epitope Prediction. *Journal of Immunology Research, 2017*, 2680160.
Sancho, A. I., Foxall, R., Rigby, N. M., Browne, T., Zuidmeer, L., van Ree, R., … Mills, E. N. C. (2006). Maturity and storage influence on the apple (Malus domestica) allergen Mal d 3, a nonspecific lipid transfer protein. *Journal of Agricultural and Food Chemistry, 54*, 5098–5104.
Schenk, M. F., Cordewener, J. H. G., America, A. H. P., Van't Westende, W. P. C., Smulders, M. J. M., & Gilissen, L. J. W. J. (2009). Characterization of PR-10 genes from eight Betula species and detection of Bet v 1 isoforms in birch pollen. *BMC Plant Biology, 9*, 24.
Schenk, M. F., Fischer, A. R., Frewer, L. J., Gilissen, L. J., Jacobsen, E., & Smulders, M. J. (2008). The influence of perceived benefits on acceptance of GM applications for allergy prevention. *Health, Risk & Society, 10*, 263–282.
Shi, Y., Jiang, L., Zhang, L., Kang, R., & Yu, Z. (2014). Dynamic changes in proteins during apple (Malus x domestica) fruit ripening and storage. *Horticulture Research, 1*, 6.
Siekierzynska, A., Piasecka-Kwiatkowska, D., Litwinczuk, W., Burzynska, M., Myszka, A., Karpinski, P., … Sozanski, T. (2021). Molecular and immunological identification of low allergenic fruits among old and new apple varieties. *International Journal of Molecular Sciences, 22*.
Son, D. Y., 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. *European Journal of Nutrition, 38*, 201–215.
Spangfort, M. D., Mirza, O., Ipsen, H., van Neerven, R. J. J., Gajhede, M., & Larsen, J. N. (2003). Dominating IgE-binding epitope of Bet v 1, the major allergen of birch pollen, characterized by X-ray crystallography and site-directed mutagenesis. *Journal of Immunology (Baltimore, Md.: 1950), 171*, 3084–3090.
Swoboda, I., Jilek, A., Ferreira, F., Engel, E., Hoffmann-Sommergruber, K., Scheiner, O., … Breitenbach, M. (1995). Isoforms of Bet v 1, the major birch pollen allergen, analyzed by liquid chromatography, mass spectrometry, and cDNA cloning. *The Journal of Biological Chemistry*, 2617–12613.
Uehara, M., Sato, K., Abe, Y., & Katagiri, M. (2001). Sequential IgE epitope analysis of a birch pollen allergen (Bet v1) and an apple allergen (Mal d1). *Allergology*. International.
Vegro, M., Eccher, G., Populin, F., Sorgato, C., Savazzini, F., Pagliarani, G., … Botton, A. (2016). Old apple (Malus domestica L. Borkh) varieties with hypoallergenic properties: An integrated approach for studying apple allergenicity. *Journal of Agricultural and Food Chemistry, 64*, 9224–9236.
Witke, W. (2004). The role of profilin complexes in cell motility and other cellular processes. *Trends in Cell Biology, 14*, 461–469. https://doi.org/10.1016/j.tcb.2004.07.003
Zaborsky, N., Brunner, M., Wallner, M., Himly, M., Karl, T., Schwarzenbacher, R., Ferreira, F., & Achatz, G. (2010). Antigen aggregation decides the fate of the allergic immune response. *Journal of Immunology (Baltimore, Md.: 1950), 184*, 725–735.
Zhang, Q., Yang, J., Bautista, J., Badithe, A., Olson, W., & Liu, Y. (2018). Epitope mapping by HDX-MS elucidates the surface coverage of antigens associated with high blocking efficiency of antibodies to birch pollen allergen. *Analytical chemistry, 90*, 11315–11323.
Zuidmeer, L., Goldhahn, K., Rona, R. J., Gislason, D., Madsen, C., Summers, C., … Keil, T. (2008). The prevalence of plant food allergies: A systematic review. *The Journal of Allergy and Clinical Immunology, 121*, 1210–1218.e4.