

Development of soybeans with low P34 allergen protein concentration for reduced allergenicity of soy-foods

Soybeans low in P34 protein for reduced allergenicity of soy-foods

Daisuke Watanabe,^a Nóra Adányi,^b Krisztina Takács,^b Anita Maczó,^b András Nagy,^b Éva Gelencsér,^b Martin Pachner,^a Kathrin Lauter,^c Sabine Baumgartner^c and Johann Vollmann^{a*}

* Correspondence to: Johann Vollmann, University of Natural Resources and Life Sciences Vienna (BOKU), Konrad Lorenz Str 24, 3430 Tulln an der Donau, Austria. E-mail: johann.vollmann@boku.ac.at

a Division of Plant Breeding, Department of Crop Sciences, University of Natural Resources and Life Sciences Vienna (BOKU), Konrad Lorenz Str 24, 3430 Tulln an der Donau, Austria

b National Agricultural Research and Innovation Centre, Food Science Research Institute (NARIC-FSRI), Herman Ottó út 15, 1022 Budapest, Hungary

c Center for Analytical Chemistry, Department IFA-Tulln, University of Natural Resources and Life Sciences Vienna (BOKU), Konrad Lorenz Str 20, 3430 Tulln an der Donau, Austria

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jsfa.7827

Abstract

BACKGROUND: In soybean, at least 16 seed proteins have been identified as causing allergic reactions in sensitive individuals. As a soybean genebank accession low in the immunodominant protein P34 (Gly m Bd 30K) has recently been found, introgression of the low-P34 trait into adapted soybean germplasm has been attempted in order to improve the safety of food products containing soybean protein. Therefore, marker assisted selection and proteomics were applied to identify and characterize low-P34 soybeans.

RESULTS: In low-P34 lines selected from a cross population, concentrations of the P34 protein as identified with a polyclonal antibody were reduced by 50-70% as compared to P34-containing controls. Using 2D-electrophoresis and immunoblotting, the reduction of P34 protein was verified in low-P34 lines. This result was confirmed by LC-MS/MS analysis, which revealed either a reduction or complete absence of the authentic P34 protein as suggested from presence or absence of a unique peptide useful for discriminating between conventional and low-P34 lines.

CONCLUSIONS: Marker-assisted selection proved useful for identifying low-P34 soybean lines for the development of hypoallergenic soy-foods. The status of the P34 protein in low-P34 lines needs further characterization. In addition, the food safety relevance of low-P34 soybeans should be tested in clinical studies.

Keywords: soybean; marker assisted selection; Gly m Bd 30K protein; food allergy; proteomics

BIOETHICS APPROVAL STATEMENT

The animal experiments described in this publication have met all the requirements of the Advisory Committee on Animal Protection for the Council of Ethics of Animal Experimentation in Hungary (official reference number: PEI/001/2734-7/2014).

INTRODUCTION

Soybean (*Glycine max* [L.] Merr.) contains about 400 g kg⁻¹ of protein in the seed, and therefore it is the major source of high-protein raw materials for both food and feed industries. Due to their favorable nutritional and functional properties, soybean ingredients such as protein, oil, hulls or lecithin are utilized in a growing number of modern foods and convenience food products. However, the presence of allergenic proteins causing abnormal immunological (i.e. immunoglobulin E mediated) reactions in sensitive subjects is limiting the use of soybean protein products,¹ as soybean might contribute to the observed increase in food allergy incidences.² In Europe, a prevalence of soybean allergy is estimated to be around 0.4% of the population based on meta-analysis results.³ For North America, up to 10% of the population are affected by food allergies,² and estimated rates of soyfood allergies of 0.4% for infants and 0.3% for adults have been reported.⁴ In Japan, about 10% of infants and 1.3-2.6% of school-aged children are considered as suffering from food allergies with 1.5% of incidences are soy allergies.⁵ For increasing food safety within the European Union, soybean is among 14 food groups requiring a specific labelling because of their allergenic potential.⁶ Moreover, immunological cross-reactivity is another concern for allergic patients, as it widens the rate of allergic incidences: So far, cross-reactivity of soybean has been identified with peanut⁷ (Xiang et al., 2002) as well as almost all other legumes,⁸ birch pollen⁹ and cow milk proteins.¹⁰

While at least 16 allergenic soybean proteins have been found,¹ the 34 kDa protein Gly m Bd 30K (P34) has been identified as immunodominant causing allergic effects in over 65% of soy-sensitive subjects.¹¹ P34 is an oil body associated protein showing sequence similarity to thiol protease of the papain family, it is present in less than 1% of total seed protein, and it preserves allergenicity even after autoclaving treatments.¹² The mature P34 protein consists of a sequence of 257 amino acids derived from a 379 amino acid precursor,¹³ and five distinct immunodominant epitopes are present.¹⁴ The P34 protein is classified as a pathogenesis-related protein of the plant defense system,¹⁵ as its biological function is that of a receptor mediating syringolide signaling in plant defense responses against bacteria by binding syringolide elicitors.¹⁶

Transgene-induced gene silencing was presented as the first approach of suppressing the expression of the P34 protein.¹⁷ Later, two natural "P34-null" genotypes were identified

expressing P34 at a low level after screening over 13,000 accessions of the USDA *Glycine max* collection.¹³ The P34 locus was reported to reside on soybean linkage group A2 (chromosome 08), and the low-P34 mutant allele was caused by a four-basepair insertion at the P34 start codon¹⁸ which has an effect on translation of the P34 gene thereby greatly reducing protein expression.¹⁹

In recent years, the uses of soybean for food production and soy-food consumption have increased due to a variety of new soy-food products developed and health benefits associated with bioactive ingredients of soybean.²⁰ For this reason, hypoallergenic soybeans containing reduced levels of allergens would be a valuable raw material for soy-food production. As the low-P34 trait is present in exotic germplasm only, the introgression of the trait into an early maturity genetic background (i.e. soybean maturity groups 000 to 0) adapted to Central European soybean growing conditions would allow for breeding and further characterization of the trait. Therefore, the objectives of the present research were i.) to select early maturity segregants from a cross between a low-P34 donor and an adapted early maturity genotype using molecular markers for identifying low-P34 lines, and ii.) to validate and characterize the low-P34 trait using immunological and proteomics methods. Thus, soybean lines could be developed for further evaluation of the low-P34 trait on its contribution to food safety.

MATERIALS AND METHODS

Genetic materials

A segregating population was developed from the cross OAC Erin × PI567476 in 2009. OAC Erin is a maturity group 000 soybean cultivar of Canadian origin. PI567476 (named “Yu ci huang”) is a maturity group IV accession collected in Shanxi, China²¹ which was obtained from the USDA genetic resources program (USDA-ARS, Urbana, IL, USA). PI567476 had previously been identified as a low P34 line (null-line) in which the P34 allergen protein was not detectable.¹³ In F₂ and F₃, the population was selected for early maturity, and 615 F_{2:3:4}-lines were genotyped by simple-sequence repeat (SSR) markers to select lines for proteomics and immunological analysis. Finally, a field experiment was carried out on 82 F_{2:3:5}-lines.

Field experimentation

Selections for early maturity, favorable pod set, lodging resistance and short plant length were carried out at the BOKU experimental station of Gross Enzersdorf near Vienna, Austria in the 2011-2013 seasons. For a preliminary evaluation of major seed quality characters, a field experiment was carried out on 82 $F_{2:3:5}$ -lines grown in 2.5 m long single row plots with 50 cm row spacing in a randomized complete block design (RCB) with two replications in the 2014 season.

SSR marker analysis

For selection of presumable low-P34 genotypes, a total of 615 $F_{2:3:4}$ -lines was genotyped by SSR markers of soybean linkage group A2 (chromosome 08), as this linkage group has been described to carry the P34 gene locus.¹⁸ Nucleotide sequences of primer pairs were utilized from previously published SSR loci.^{22,23} Soybean DNA extraction as well as amplification and evaluation of PCR products has been described elsewhere.²⁴ A linkage analysis between SSR markers was carried out using the MAPMAKER/EXP 3.0 program.²⁵

Development of polyclonal antibodies specific to P34 protein

The antibodies were developed using two out-bred New Zealand White rabbits weighing 2–2.5 kg for the immunization with recombinant P34 (rP34, >95% purity) soybean protein (Biorbyt, Cambridge, UK).²⁶ For the primary injection (subcutaneous) of rabbits 25 μ g antigen kg^{-1} body weight emulsified in Freund's Complete Adjuvant (FCA) (1:1) was used, which was followed by 4 booster injections (subcutaneous) in 14 day intervals (days: 14, 28, 42, 56) with 25 μ g antigen kg^{-1} body weight emulsified in Freund's Incomplete Adjuvant (FIA) (1:1). Rabbit antibodies (RbIgG) specific to rP34 derived from hyperimmune serum were obtained by 70% ammonium-sulphate (pH 7.4) precipitation (threefold), dialyzed and freeze-dried. The compliance of serum-titer was controlled by indirect ELISA during the procedure as well as after serum purification. The adequately chosen serum dilution was used for the development of rP34-specific competitive indirect ELISA.

Competitive indirect ELISA

Total protein was isolated from 0.5 g of finely powdered soybean seed samples using 1.5 ml of extraction buffer containing 0.15 M PBS and 0.05% Tween-20 (pH 7.4) by extracting for 30 min

and centrifuging at $8,000 \times g$ for 15 min; three layers (i.e. insoluble rest, buffer, oil body pad) were separated. The intermediate buffer layer was collected into tubes to be used for ELISA. A competitive indirect ELISA assay was performed.²⁷ A microtiter plate was coated with $5 \mu\text{g ml}^{-1}$ of rP34 soy protein (Biorbyt, Cat.no. orb81523) in 0.05 M carbonate-bicarbonate buffer (pH 9.6) followed by an overnight incubation at 4°C . After incubation, the free binding sites were blocked with 0.2 ml/well 0.15 M PBS (pH 7.4) containing 0.5 % gelatin and incubated with shaking for 1 hour at 37°C . Thereafter, 0.05 ml/well rP34 protein standards (10^3 - $10^{-3} \mu\text{g ml}^{-1}$ diluted in 0.15 M PBS) or extracted soy proteins from samples were added to each well followed by 0.05 ml/well anti-recombinant P34 rabbit IgG antibody (a-rP34 Rb IgG) (1:1250, diluted in PBS) and incubated for 1 hour at 37°C . Afterwards, 0.1 ml/well a-Rb IgG, HRP conjugate (Sigma, St. Louis, MO, USA, 1:20,000 diluted in 0.15 M PBS) was added to the immune-complex and shaken for 1 hour at 37°C . After the incubation, 0.1 ml of OPD- H_2O_2 reagent (Sigma) was added (no shaking for 10 min under dark conditions at room temperature). The reaction was stopped with 0.05 ml of 2M H_2SO_4 , and the absorbance was measured at 492/630 nm using an ELISA plate reader (Dynatech, MR 7000, Dynatech Labs, Chantilly, VA, USA). Between the incubation steps, the plate was washed out with 0.15 M PBS-0.1% Tween-20 (pH 7.4), and then the water-drops were removed. The concentration of P34 protein in each sample was calculated using the linear part of the calibration curve ($R^2 = 0.992$, $y = -0.107 \ln(x) + 0.711$) of rP34 protein standards and given in $\mu\text{g P34 protein ml}^{-1}$. Based on the calibration curve, the recovery range of the indirect ELISA was 79.3-103.3% (average (102.7%)) for four measurement points with two parallel measurements. Results presented later on are given in $\mu\text{g P34 protein mg}^{-1}$ of powdered soybean sample (see Fig. 2).

Protein separation by 2D SDS-PAGE

Standard two-dimensional protein electrophoresis^{28,29} was performed: For the first dimension 7 cm IPG (immobilized pH gradient) strips with pH range of 3-10 were used. The IPG strips were rehydrated in 250 μl rehydration buffer (8 M urea, 1% CHAPS [3-(3- cholamidopropyl-dimethylammonio)-1-propanesulfonate hydrate], 20 mM DTT [DL-threo-1,4-dimercapto-2,3-butanediol]) containing 800 μg of soybean meal. The isoelectric focusing (IEF) was carried out by linear increasing voltage from 250 V to 24.000 Vh using 7 cm IPG (immobilized pH gradient) strip with pH range of 3-10 in a PROTEAN IEF cell (Bio-Rad, Hercules, CA, USA). Following

isoelectric focusing the strips were incubated in equilibrating buffer (2% SDS, 6 M urea, 1.5 M Tris-HCl (pH 8.8), 20% glycerol) containing 2% DTT for 15 min, and 2.5% iodoacetamide for 15 min at room temperature. For the second dimension, electrophoresis was performed on 15% SDS-PAGE for 1 h at 200 V using a Mini-PROTEAN II cell (Bio-Rad). Subsequently, the gels were fixed in 20% TCA (trichloroacetic acid) (for 20 min) and stained with Coomassie Brilliant Blue R-250.

Western blot

For western blotting, the 2D-separated proteins were transferred from the gel onto polyvinylidene fluoride membrane (0.45 μm , Millipore Corp., Billerica, MA, USA) by semi-dry blotting (Bio-Rad Trans Blot SD Semi-Dry Transfer Cell) for 60 min at 0.25 V, 0.08 mA cm^{-2} . The transferred proteins were submitted overnight under reaction with α -rP34 Rb IgG serum (1:200, dark condition at room temperature). After that, α -Rb IgG, HRP conjugate (1:600) was added to the membrane for 1 hour incubation with shaking at room temperature. Subsequently, the membrane was treated with substrate solution containing 4-chloronaphthol (Sigma), H_2O_2 and 0.15 M PBS (pH 7.4) in ethanol. Between the incubation steps, the membrane was washed out from the unnecessary reagents with 0.15 M PBS containing 0.05% Tween 20. Images were scanned by a Bio-Rad Gel Doc 2000 system.

Protein extraction and trypsin digestion

In protein extraction for tryptic digestion and LC-MS/MS analysis,^{30,31} an amount of 500 mg of powdered soybean was mixed with 5 ml Tris-Borat-buffer (8mM Tris, 10 mM $(\text{NH}_4)_2\text{B}_{10}\text{O}_{16}$, pH 8.5). The extraction was incubated in a water bath at 37°C for 2 h without mixing. The mixture was centrifuged for 30 min at $9,299 \times g$, 4°C using an ALLEGRA® X-22R centrifuge (Beckman Coulter, Inc., Fullerton, CA, USA). The clear middle layers were collected into new tubes. Protein contents were evaluated with 2D-Quant Kit (GE-Healthcare, Little Chalfont, UK), and samples were set to a final concentration of 83.3 $\mu\text{g ml}^{-1}$ of protein in 50 mM ABC₅₀. 120 μl of extracted protein solution were mixed with 50 μl of 10 mM 1,4-dithiothreitol (DTT) and incubated at 60°C for 1h. Afterwards, 50 μl of 55 mM iodoacetamid (IAA) were added to the mixture and incubated at 25°C under dark conditions for 1h. Subsequently, 16 μl of trypsin solution, resulting in a trypsin to protein ratio of 1:5, were mixed with the samples and incubated at 37°C

overnight. The reaction was stopped by 26 μ l of 5% formic acid.

Liquid chromatography-tandem mass spectrometry

The final samples were centrifuged for 15 min at $9,299 \times g$ at room temperature, and the clear supernatant was transferred into the MS-vials before the detection. The chromatographic separation of tryptically-digested samples was carried out on an Aeris Peptide XB-C18 (150 $\times$ 2.1 mm, 3.6 μ m particle size from Phenomenex, Torrance, CA, USA), equipped with a C18 4 $\times$ 3 mm i.d. security guard cartridge (Phenomenex) using an Agilent 1290 Infinity LC (Agilent Technologies, Santa Clara, CA, USA). Injection volume was 15 μ l, and the flow rate was 0.3 μ l min⁻¹. Eluent A consisted of water and 0.1% formic acid, and eluent B consisted of methanol and 0.1% formic acid. The HPLC separation was performed with a 15 minutes-gradient from 5% to 70% B.

For mass spectrometry analysis, a QTrap 6500 LC-MS/MS system (Applied Biosystems, Foster City, CA, USA) with an IonDriveTMTurbo V source was used. Chromatographic and (tandem) mass spectrometric data were analyzed using the Skyline software.³² The electrospray ionisation (ESI) source parameters used were defined for positive ionisation mode: curtain gas, 40 psi; both ion source gas 1 and 2, 50 psi; source temperature, 400°C; and +4,000 V of ion spray voltage. The collision gas (nitrogen) was defined to high. The pause between mass ranges was defined to 5 ms. *In silico* tryptic digestion of the P34 protein for marker peptide identification was performed using the web-based software tool PeptideMass³³ and the UniProt³⁴ protein database.

Analysis of seed characters

Soybean total seed protein and oil content from the 2014 field experiment were determined from ground seed samples by Fourier-transform near infrared reflectance spectroscopy (FT-NIRS) using a Bruker Matrix-I NIRS instrument (Bruker, Ettlingen, Germany) and calibrations described elsewhere.³⁵

Statistical analysis

For statistical analysis of data, Student's *t*-test or *F*-test following analysis of variance were used, and *P* values < 0.05 were considered as statistically significant.

RESULTS

Marker assisted selection

Out of 48 SSR markers assigned to soybean linkage group A2, 23 were polymorphic between the two parents, and 13 of them produced clearly separable bands between the parental alleles and therefore were used to analyze the 615 progeny lines. Based on a linkage map calculated (Fig. 1) and the assumption of close linkage between SSR marker Sat_215 and the P34 locus, presumable low-P34 genotypes were selected for further analysis if they were homozygous for the PI567476 parental allele at the Sat_215 marker locus as well as at the two linked loci Satt632 and Satt424. Genotypes homozygous for the OAC Erin allele at all three loci were selected as conventional P34 lines for control purposes.

P34 protein concentration by competitive indirect ELISA

Seven lines carrying the low-P34 marker, one conventional P34 line and the two parental genotypes were subject to rP34-specific competitive indirect ELISA to determine the concentration of the P34 protein as detected by the polyclonal antibody (Fig. 2). Six out of seven low-P34 lines showed a significantly lower P34 protein concentration (i.e. 50-70% reduction) as compared to the controls. In contrast, the P34 protein concentration of the low-P34 donor PI567476 was not different from that of OAC Erin; however, PI567476 was grown in a greenhouse environment whereas all other samples used in Fig. 2 were from a harvest of the same field environment.

Two-dimensional SDS-PAGE and western blotting

In order to visualize the presence or absence of the P34 protein, both parents as well as the low-P34 line GT9X-13-1 (selection based on SSR marker analysis) were subject to two-dimensional electrophoresis and subsequent immunoblotting. As a positive control of the procedure and for testing the polyclonal antibody used, the pure recombinant P34 (rP34) protein yielded several spots at approximately 34 kDa molecular weight and the reported pH range of around pH 4.5 (Fig. 3). In Fig 4, all separated proteins stained (a, c, e) or western blots for immunodetection of the P34 protein (b, d, f) are presented for parents OAC Erin (a, b) and PI567476 (c, d) as well as GT9X-13-1 (e, f), respectively. While the P34 protein spots are clearly present in OAC Erin (Fig. 4b), they are neither detected in the immunoblots of PI567476 (Fig.

4d) nor of line GT9X-13-1 (Fig. 4f). Likewise, a spot located at the 34 kDa position in Fig. 4a (OAC Erin) is lacking on the gels given in Figs. 4c and 4e for PI567476 and GT9X-13-1, respectively.

Soybean P34 detection by LC-MS/MS analysis

In silico tryptic digestion of the soybean P34 protein resulted in six peptides useful for peptide identification by LC-MS/MS. Three of these peptides were used as marker peptides as they are unique to the authentic soybean P34 protein according to the UniProt database (Table 1); for comparison, the respective peptide sequences of a P34-pseudogene are given as well. After sample digestion with trypsin, six different soybean samples were scanned by MS/MS and three peptide peaks were identified corresponding to the marker peaks (Fig. 5). Total ion chromatograms of low-P34 lines (Fig. 5 a, b, c) and the low-P34 parent (Fig. 5 e) show peptide peaks 1 and 2 only, and peak 2 is expressed at a much lower level than in conventional genotypes (Fig. 5 d and f). In contrast, the third peptide peak (peak 3) is only detected in chromatograms of a P34-containing line (Fig. 5 d) as well as in the conventional parent OAC Erin (Fig. 5 f) at retention time of 6.1 min. Thus, peptide 3 can serve as an unambiguous marker indicating the presence or absence of the P34 protein.

Seed quality

In a replicated field experiment with $F_{2:3:5}$ -lines, no statistically significant difference between 43 low-P34 lines and 14 conventional P34 lines was found for total seed protein and oil content, whereas low-P34 lines had a slightly lower thousand seed weight as compared to conventional lines (F -test, $P = 0.0426$, detail data not shown).

DISCUSSION

The identification and genetic characterization of a low-P34 soybean^{13,18} which is controlled by a single gene,³⁶ for the first time opens up an option to reduce a major soybean allergen through a plant breeding approach. As soybean protein is a widely utilized raw material in the food industry, hypoallergenic soybean protein would be of considerable interest for increasing food safety of a large number of food products.

The use of immunological or proteomics methods in screening for low-P34 segregants is both costly and laborious and is therefore not a suitable procedure for analyzing large numbers of samples in selection programs. In contrast, polymorphic SSR markers linked to the P34 locus (Fig. 1) proved to be efficient in identifying low-P34 segregants. While there is little sequence difference between low-P34 and normal soybeans in the coding region of the P34 gene resulting in four amino acid changes between the protein sequences only,¹³ clear evidence has been presented that a 4-basepair insertion within the promoter region of the gene upstream of the start codon is the major factor for down-regulation of the P34 protein in low-P34 lines on the level of translation.¹⁹ This finding bears options for the development of DNA-based selection markers,³⁷ and a U.S. patent for marker development utilizing the variation in promoter sequence has also been granted.³⁸ Alternatively, polymorphic SSR markers such as Satt632, Sat_215 and Satt424 as utilized in the present study or a number of tightly linked SSRs could be applied in selection, which would also ensure that most of the linkage group stems from the adapted parent, as recombination events could be visualized.^{22,23}

In six out of seven low-P34 lines, a significant reduction of P34 concentration has been found by competitive indirect ELISA in the present study (Fig. 2). Under the conditions described, small amounts of P34 protein (0.03-0.07 $\mu\text{g mg}^{-1}$ powdered sample) were detected by the polyclonal antibody demonstrating the reduction in P34 protein concentration through the low-P34 allele. However, the amounts of P34 are probably overestimated because of the possibility of unspecific binding of the antibody to suitable epitopes of other proteins present. Unexpectedly, the amounts of P34 protein were high in the low-P34 parent sample (PI567476) which had been grown in a different environment (i.e. greenhouse because of late maturity) as compared to all other genotypes (see Fig. 2). This is probably due to a relative overexpression of P34 or similar proteins in the absence of biological nitrogen fixation or to differences in protein extraction efficiency as also shown for P34 at different developmental stages.³⁹ ELISA results were supplemented and confirmed by immunoblot analyses, in which the sensitivity of the polyclonal antibody utilized was clearly proven by western blots of recombinant P34 protein at about 34 kDa and pH 4.5 (Fig. 3). Different additional spots (25 kDa and 37 kDa) may be either contaminations or due to protein denaturing and protein isoforms with variable isoelectric point of the P34 protein, as also shown elsewhere.^{17,39} An eight-fold reduction of the P34 protein level has been estimated by western blotting for genotypes carrying the low-P34 allele as

compared to conventional P34 protein expression.¹⁸ The present immunological proteome analysis of conventional vs. low-P34 samples showed clear-cut differences between genotypes (Fig. 4): P34 spots were totally absent in western blots of low-P34 samples (Figs 4d, 4f) as compared to the conventional parent OAC Erin (Fig. 4b). In addition, a number of further protein spots detected by the polyclonal antibody in all three western blots are probably caused by cross-reactivity of the rP34 antibody with other proteins. Opposite to our results, immune-reactive spots of P34 were present at reduced levels in other studies.^{13,18,39}

From tryptic digestion and tandem mass spectrometric identification of the P34 protein, a peptide was identified to serve as a marker (peptide 3 in Fig. 5 and Table 1) for discriminating between low and high P34 lines. While peptide 1, which represents a sequence close to the starting region of the protein (amino acids 39-48) is detectable in all samples, peptide 3 (amino acids 358-367) is only found in high P34 lines (Fig. 5). This finding raises the question of protein expression in low-P34 lines: Based on sequenced cDNA of the low-P34 accession, a low-P34 protein has been predicted differing in four amino acids only from a wildtype (cv. Century) P34 protein;¹³ this sequence difference described would not affect the detection of the marker peptides in Fig. 7, as the respective sequences are not affected. Moreover, an even identical cDNA sequence between low-P34 genotype PI567476 and cv. Clark was described,¹⁹ while it was demonstrated that the insertion in the promoter region strongly affects translation efficiency being the reason for the low-P34 trait. This is also supported by the present results which reveal that peptide 1 is translated in all genotypes, whereas peptide 3 is absent in low-P34 lines (while peptide 2 is behaving in an intermediate manner, see Fig. 5) suggesting a truncation of the protein between peptides 1 and 3 in low-P34 lines. Although P34-pseudogenes¹⁸ present in the paleopolyploid genome of soybean are most probably not expressed, peptide 1 could also reveal expression of a pseudogene protein due to sequence identity with the authentic P34 peptide (Table 1), whereas the predicted pseudogene amino acid sequences differ from that of the authentic gene in peptides 2 and 3 and would therefore not be detected in low-P34 lines. Furthermore, two pseudogenes contain stop codons after either 59 or 142 amino acids,¹⁸ therefore they both would synthesize peptide 1 only before protein truncation. Pseudogene expression would not be relevant with respect to allergenic potential, as the five P34-epitopes¹⁴ do all have a different amino acid sequence in the authentic P34 protein as compared to the predicted sequence at the Glyma08g-12280.1 pseudogene. The expression of the P34 protein

in low-protein lines derived from PI567476 requires further investigation: While LC-MS/MS-based results suggest the expression of a truncated protein (Fig. 7) which could be due to posttranslational modifications,¹⁶ other findings reported the presence of the authentic P34 protein at reduced concentrations^{18,39} which might be low enough to bypass the threshold level for allergenicity.⁴⁰

As the immunological basis of soybean allergy has been described as highly individual and complex due to the reaction against different allergens as well as individual dose depending symptom development,⁴¹ the low-P34 soybean material could now be evaluated for its relevance in reducing allergic reactions in soybean sensitive subjects. Subsequently, it could be introgressed to food-grade soybean cultivars for the development of low-allergenic germplasm. Recently, a “Triple Null” soybean has been produced by backcrossing the low-P34, the low Kunitz trypsin inhibitor and the null soybean agglutinin alleles into the cultivar Williams 82.⁴⁰ They reported a reduction of total seed protein content in the “Triple Null” genotype which could be due to an uncompensated loss of three proteins. As no significant differences in protein or oil content between high and low P34 lines have been measured in the present study, the decrease in protein content found⁴⁰ might primarily be caused by the reduction in the Kunitz trypsin inhibitor protein, as observed in other populations as well.⁴² Nevertheless, stacking of low-P34 and other food safety characters⁴⁰ is highly relevant for improving the value of soybean protein. In the present study, both parent genotypes share a low cadmium accumulation trait²⁴ in addition to segregation in P34, which gives a similar stack of desirable traits to soybean breeding and for the subsequent development of soybean products with improved food safety.

REFERENCES

- 1 L'Hocine L and Boye JI, Allergenicity of soybean: New developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. *Crit Rev Food Sci Nutr* **47**:127–143 (2007).
- 2 Sicherer SH, Epidemiology of food allergy. *J Allergy Clin Immunol* **127**:594–602 (2011).
- 3 Nwaru BI, Hickstein L, Panesar SS, Roberts G, Muraro A and Sheikh A on behalf of the EAACI Food Allergy and Anaphylaxis Guidelines Group, Prevalence of common food allergies in Europe: a systematic review and meta-analysis. *Allergy* **69**:992–1007 (2014).

- 4 Sicherer SH and Sampson HA, Food allergy. *J Allergy Clin Immunol* **125**:S116–S125 (2010).
- 5 Imai T, Syokubutsu allergy no eiyou shidou no tebiki 2011 (Nutritional guidance about Food allergy 2011). *Ministry of Health, Labour, Welfare Japan*, 1–22 (2011).
- 6 European Commission, Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011. *Off J Eur Union* L 304/18 (2011).
- 7 Xiang P, Beardslee TA, Zeece MG, Markwell J and Sarath G, Identification and analysis of a conserved immunoglobulin E-binding epitope in soybean G1a and G2a and peanut Ara h 3 glycinins. *Arch Biochem Biophys* **408**:51–57 (2002).
- 8 Martinez San Ireneo M, Ibanez Sandin MD and Fernandez-Caldas, E, Hypersensitivity to members of the botanical order Fabales (legumes). *J Investig Allergol Clin Immunol* **10**:187–199 (2000).
- 9 Mittag D, Vieths S, Vogel L, Becker W-M, Rihs H-P, Helbling A, Wüthrich B and Ballmer-Weber BK, Soybean allergy in patients allergic to birch pollen: clinical investigation and molecular characterization of allergens. *J Allergy Clin Immunol* **113**:148–154 (2004).
- 10 Kattan JD, Cocco RR and Järvinen KM, Milk and soy allergy. *Pediatr Clin North Am* **58**:407–426 (2011).
- 11 Ogawa T, Bando N, Tsuji H, Okajima H, Nishikawa K and Sasaoka K, Investigation of the IgE-binding proteins in soybeans by immunoblotting with the sera of the soybean-sensitive patients with atopic dermatitis. *J Nutr Sci Vitaminol* **37**:555–565 (1991).
- 12 Yamanishi R, Huang T, Tsuji H, Bando N and Ogawa T, Reduction of the soybean allergenicity by the fermentation with *Bacillus natto*. *Food Sci Technol Int* **1**:14–17 (1995).
- 13 Joseph LM, Hymowitz T, Schmidt MA and Herman EM, Evaluation of *Glycine* germplasm for nulls of the immunodominant allergen P34/Gly m Bd 30k. *Crop Sci* **46**:1755–1763 (2006).
- 14 Helm RM, Cockrell G, Connaughton C, West CM, Herman E, Sampson HA, Bannon GA and Burks AW, Mutational analysis of the IgE-binding epitopes of P34/Gly m Bd 30K. *J Allergy Clin Immunol* **105**:378–384 (2000).
- 15 Breiteneder H and Radauer C, A classification of plant food allergens. *J Allergy Clin Immunol* **113**:821–830 (2004).
- 16 Ji C, Boyd C, Slaymaker D, Okinaka Y, Takeuchi Y, Midland SL, Sims JJ, Herman E and

- Keen N, Characterization of a 34-kDa soybean binding protein for the syringolide elicitors. *Proc Natl Acad Sci USA* **95**:3306–3311 (1998).
- 17 Herman EM, Helm RM, Jung R and Kinney AJ, Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol* **132**:36–43 (2003).
- 18 Bilyeu K, Ren C, Nguyen HT, Herman W and Sleper DA, Association of a four-basepair insertion in the P34 gene with the low-allergen trait in soybean. *Plant Genome* **2**:141–148 (2009).
- 19 Koo SC, Seo JS, Park MJ, Cho HM, Park MS, Choi CW, Jung W-H, Lee KH, Jin BJ, Kim SH, Shim SI, Chung J-S, Chung JI and Kim MC, Identification of molecular mechanism controlling *P34* gene expression in soybean. *Plant Biotechnol Rep* **7**:331–338 (2013).
- 20 Chen K-I, Erh M-H, Su N-W, Liu W-H, Chou C-C and Cheng K-C, Soyfoods and soybean products: from traditional use to modern applications. *Appl Microbiol Biotechnol* **96**:9–22 (2012).
- 21 Li Z and Nelson RL, Genetic diversity among soybean accessions from three countries measured by RAPDs. *Crop Sci* **41**:1337–1347 (2001).
- 22 Hwang T-Y, Sayama T, Takahashi M, Takada Y, Nakamoto Y, Funatsuki H, Hisano H, Sasamoto S, Sato S, Tabata S, Kono I, Hoshi M, Hanawa M, Yano C, Xia Z, Harada K, Kitamura K and Ishimoto M, High-density linkage map based on SSR markers in soybean. *DNA Res* **16**:213–225 (2009).
- 23 Song Q, Jia G, Zhu Y, Grant D, Nelson RT, Hwang E-Y, Hyten DL and Cregan PB, Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. *Crop Sci* **50**:1950–1960 (2010).
- 24 Vollmann J, Lošák T, Pachner M, Watanabe D, Musilová L and Hlušek J, Soybean seed cadmium concentration: validation of a QTL affecting seed cadmium accumulation for improved food safety. *Euphytica* **203**:177–184 (2015).
- 25 Lincoln S, Daly M and Lander E, Constructing genetic maps with MAPMAKER/EXP 3.0. Whitehead Institute Technical Report, 3rd edn (1992).
- 26 Harboe N and Ingild A, Immunization, isolation of immunoglobulins, estimation of antibody titre. *Scand J Immunol* **2** (Suppl s1):161–164 (1973).
- 27 Tijssen P, *Practice and theory of enzyme immunoassays*. Series: *Laboratory techniques in biochemistry and molecular biology*. Vol. 15, Series ed. by Burdon RH and van

- Knippenberg PH. Elsevier, Amsterdam, The Netherlands (1985).
- 28 Laemmli UK, Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685 (1970).
- 29 Hajós G, Gelencser E, Pusztai A, Grant G, Sakhri M, Bardocz S, Rowett T and Aberdeen AB, Biological effects and survival of trypsin inhibitors and the agglutinin from soybean in the small intestine of the rat. *J Agric Food Chem* **43**:165–170 (1995).
- 30 Ansari P, Stoppacher N and Baumgartner S, Marker peptide selection for the determination of hazelnut by LC-MS/MS and occurrence in other nuts. *Anal Bioanal Chem* **402**:2607–2615 (2012).
- 31 Costa J, Ansari P, Mafra I, Oliveira MBPP and Baumgartner S, Assessing hazelnut allergens by protein- and DNA-based approaches: LC-MS/MS, ELISA and real-time PCR. *Anal Bioanal Chem* **406**:2581–2590 (2014).
- 32 MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC and MacCoss MJ, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**:966–968 (2010).
- 33 PeptideMass, Available at http://web.expasy.org/peptide_mass, accessed 17 March 2015.
- 34 UniProt, Uniprot database. Available at <http://www.uniprot.org/uniprot>, accessed 17 March 2015.
- 35 Vollmann J, Walter H, Sato T and Schweiger P, Digital image analysis and chlorophyll metering for phenotyping the effects of nodulation in soybean. *Comp Electron Agric* **75**:190–195 (2011).
- 36 Sung MK, Seo JS, Kim KR, Han EH, Nam JW, Kang DS, Jung WS, Kim MC, Shim SI, Kim KM and Chung JI, Inheritance of P34 allergen protein in mature soybean seed. *Kor J Breed Sci* **43**:115–119 (2011).
- 37 Jeong K-H, Choi M-S, Lee S-K, Seo M-J, Hwang T-Y, Yun H-T, Kim H-S, Kim J-T, Kwon Y-U and Kim Y-H, Development of low Gly m Bd 30K (P34) allergen breeding lines using molecular marker in soybean [*Glycine max* (L.) Merr.]. *Plant Breed Biotech* **1**:298–306 (2013).
- 38 Bilyeu KD, Development of low allergen soybean seeds using molecular markers for the P34 allele. US Patent No 7,951,537 B2. (2011).
- 39 Koo SC, Bae DW, Seo JS, Park KM, Choi MS, Kim SH, Shim SI, Kim KM, Chung JI and Kim

MC, Proteomic analysis of seed storage proteins in low allergenic soybean accession. *J Korean Soc Appl Biol Chem* **54**:332–339 (2011).

40 Schmidt MA, Hymowitz T and Herman EM, Breeding and characterization of soybean *Triple Null*; a stack of recessive alleles of Kunitz trypsin inhibitor, soybean agglutinin, and P34 allergen nulls. *Plant Breed* DOI: 10.1111/pbr.12265 (2015).

41 Ballmer-Weber BK, Holzhauser T, Scibilia J, Mittag D, Zisa G, Ortolani C, Oesterballe M, Poulsen LK, Vieths S and Bindslev-Jensen C, Clinical characteristics of soybean allergy in Europe: A double-blind, placebo-controlled food challenge study. *J Allergy Clin Immunol* **119**:1489–1496 (2007).

42 Vollmann J, Schausberger H, Bistrich H and Lelley T, The presence or absence of the soybean Kunitz trypsin inhibitor as a quantitative trait locus for seed protein content. *Plant Breed* **121**:272–274 (2002).

Table 1.

Tryptic peptides (in silico digestion) of three unique P34 peptide sequences (UniProt database, accession no. O64458 [P34]) from the authentic P34 protein and the comparative amino acid sequences of a P34 pseudogene (Glyma08g 12280.1, sequences from Phytozome; <http://www.phytozome.net>)

peptide	gene	peptide		predicted	
		amino acid sequence of tryptic peptide	position in protein	molecular weight (Da)	retention time (min)
1	Authentic P34	QVSSLFQLWK	39 – 48	1235.7	7.7
	Pseudogene	QVSSLFQLWK			
2	Authentic P34	NSWGEDWGEDGYIWIQR	321 – 337	2110.9	12.5
	Pseudogene	NSFGKDWGMDGYIWIQR			
3	Authentic P34	EESETLV SAR	358 - 367	1120.5	6.1
	Pseudogene	EKSETLSGLL			

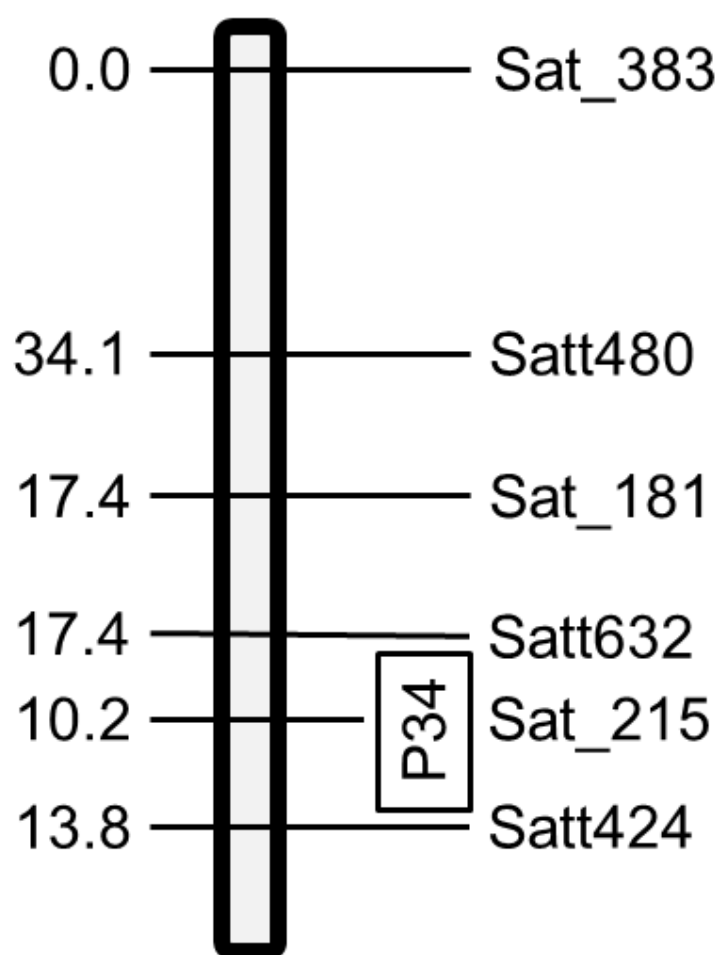


Figure 1. Genetic linkage map of soybean chromosome 08 (linkage group A2) from the cross OAC Erin × PI567476 using 6 linked SSR markers (distances in centi Morgans) and the approximate position of the P34 gene locus.

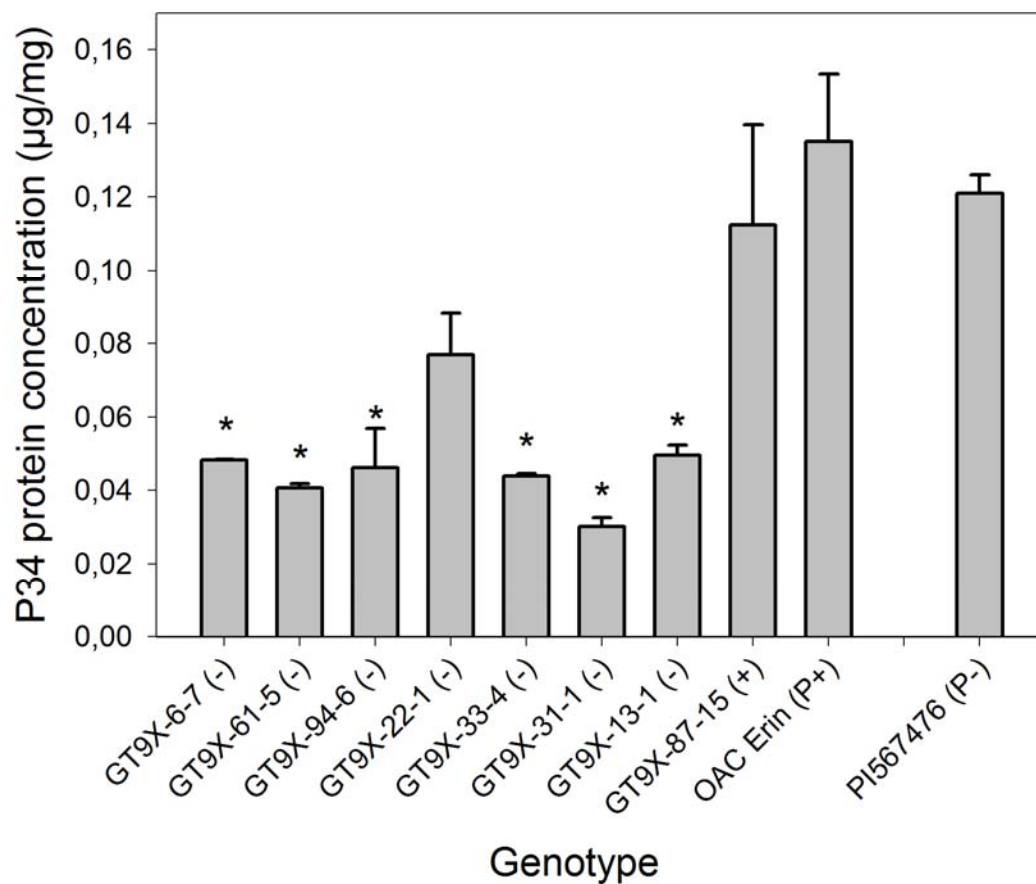


Figure 2. P34 protein concentration of soybean low-P34 lines (-) vs. conventional lines (+) using competitive indirect ELISA. The protein concentrations are given as means $\pm$ SE (bar); *: Significance ($P < 0.05$) of unpaired t -test of low-P34 lines versus OAC Erin and line GT9X-87-15.

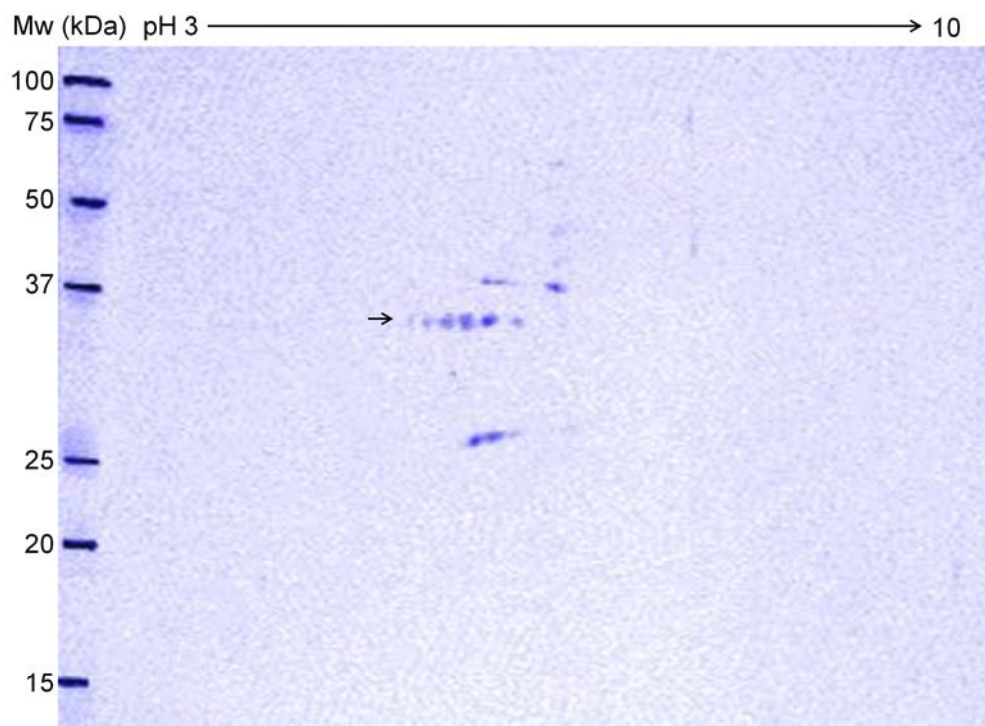


Figure 3. Control immunoblot of the recombinant P34 protein using a polyclonal P34 antibody.

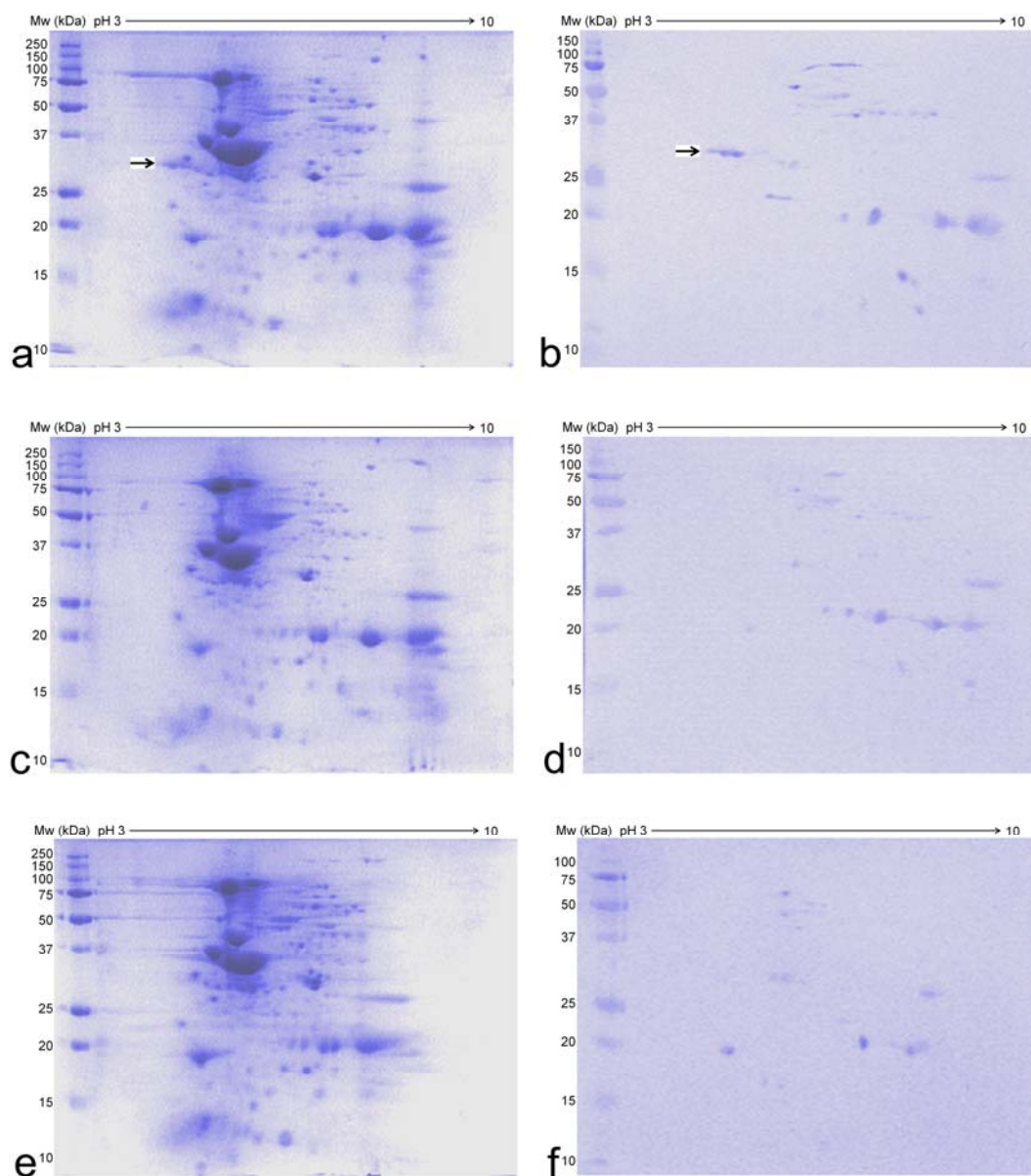


Figure 4. Two-dimensional protein electrophoresis (a, c, e) and the corresponding immunoblots (b, d, f) of soybean cv. OAC Erin (a, b; conventional parent with P34 protein expression indicated by arrows in a and b), genotype PI567476 (c, d; low P34 parent) and the low-P34 breeding line GT9X-13-1 (e, f; P34 protein spots absent).

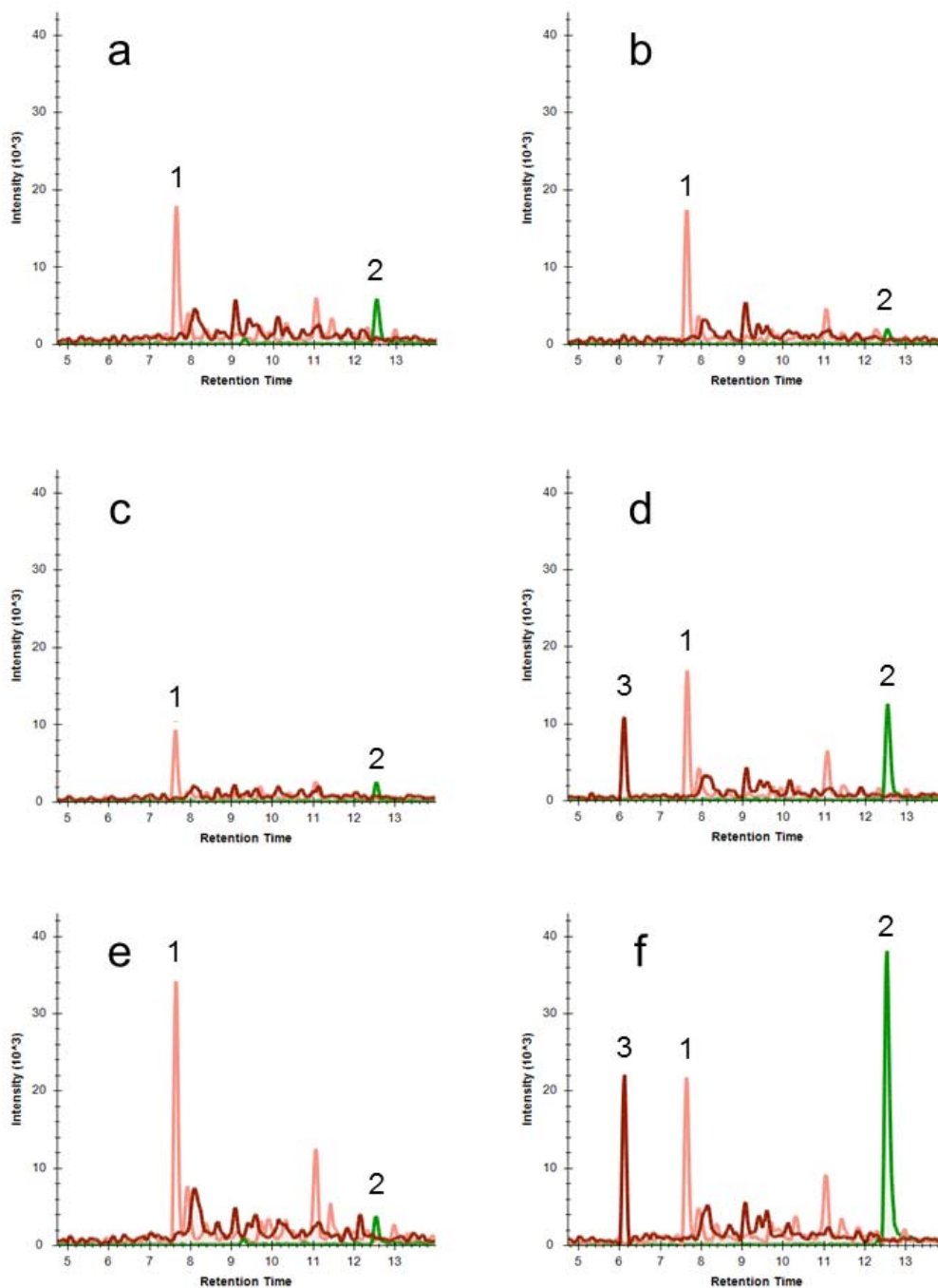


Figure 5. Total ion chromatograms of 6 different soybean genotypes for detection of three marker peptides (peaks 1, 2 and 3) from a tryptic digestion of the P34 protein. Low-P34 lines GT9X-61-5 (a), GT9X-31-1 (b) and GT9X-13-1 (c), P34-containing line GT9X-87-15 (d), low-P34 parent genotype PI567476 (e) and conventional parent OAC Erin (f). *Peak 1:* QVSSLFQLWK, *Peak 2:* NSWGEDWGEGDYIWIQR, *Peak 3:* EESETLVSAR.