

Association of a Four-Basepair Insertion in the P34 Gene with the Low-Allergen Trait in Soybean

Kristin Bilyeu,* Chengwei Ren, Henry T. Nguyen, Eliot Herman, and David A. Sleper

Abstract

The value of soybean [*Glycine max* (L.) Merr.] is determined in part by functionality of major seed components, extractable oil and high protein meal. Soybean meal is an important protein source for animal feeds, and use of soybean meal by the food industry is increasing. Food allergens have negative impacts for both foods and feeds. The major soybean allergen, Gly m Bd 30K, was identified as the P34 protein, a low abundance but highly conserved seed protein. Recently, two soybean germplasm accessions (PI 567476 and PI 603570A) were identified as having reduced P34 content without other collateral changes in seed protein composition. In order to develop molecular markers to enhance breeding for low P34, our objective was to characterize the P34 gene region in wild-type and low-P34 soybean lines and exploit any sequence differences to allow direct selection for mutant P34 alleles. The results indicated the two low-P34 germplasm accessions contain a four-basepair insertion at the P34 start codon. The mutant allele may inhibit efficient translation initiation of the P34 protein. Inheritance of the mutant P34 alleles completely associated with the low-P34 phenotype. Molecular markers were developed that were based on the mutant allele of the P34 gene.

SOYBEAN [*Glycine max* (L.) Merr.] is an important source of vegetable oil and high protein meal. Antinutritional compounds such as phytate, raffinose family oligosaccharides, and allergenic proteins that accumulate during normal soybean seed development limit the extent to which soybean meal can be included in livestock diet formulations. In addition, the United States Food and Drug Administration recently recognized soybean as one of the major food allergens and new food allergen labeling requirements are in effect.

Several soybean proteins have been identified as allergens; the major allergen, Gly m Bd 30K, provoked a response in almost two thirds of patients examined in one study (Ogawa et al., 1991). The Gly m Bd 30K protein was identified as P34, a papain superfamily cysteine proteinase type of enzyme (Kalinski et al., 1990). P34 is not an abundant seed protein, but it is consistently present in nearly all germplasm accessions evaluated (Joseph et al., 2006; Xu et al., 2007; Yaklich et al., 1999).

Transgenic suppression was successful in downregulation of P34 accumulation in soybean seeds with no apparent negative consequences to normal seed development and composition (Herman et al., 2003). While the P34 protein failed to accumulate in transgenic seeds, there were no collateral alterations in other seed proteins as assessed by 2D gel electrophoresis. Additionally, the protein storage vacuoles (PSVs) in the P34 suppressed seeds were indistinguishable from those in control seeds

Published in The Plant Genome 2:141–148. Published 10 July 2009.
doi: 10.3835/plantgenome2009.01.0006
© Crop Science Society of America
677 S. Segoe Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

K.D. Bilyeu, USDA-ARS, Univ. of Missouri-Columbia, 108 Waters Hall, Columbia, MO 65211; C. Ren, Beta Seed, Inc., 34303 Hwy. 99E, Tangent, OR 97389; H. Nguyen, Division of Plant Sciences, Univ. of Missouri-Columbia, 25 Agriculture Bldg, Columbia, MO 65211; E. Herman, USDA-ARS, D. Danforth Plant Science Center, 975 N. Warson Rd., St. Louis, MO 63132; and D.A. Sleper, Division of Plant Sciences, Univ. of Missouri-Columbia, Life Sciences Center, Columbia, MO 65211. Received 19 Jan. 2009.
*Corresponding author (bilyeuk@missouri.edu).

Abbreviations: EST, expressed sequence tags; PCR, polymerase chain reaction; PSV, protein storage vacuoles.

despite the fact that the P34 protein has been shown to accumulate in the PSVs (Herman et al., 2003).

Of approximately 16,000 soybean accessions from the USDA germplasm collections screened, only two *Glycine max* lines were found to have reduced P34 protein accumulation in seeds (Joseph et al., 2006). *Glycine max* low-P34 soybean accessions, PI 567476 and PI 603570A, were characterized as having normal accumulation of seed proteins other than P34 (Joseph et al., 2006). The objective of this work was to characterize the molecular genetic basis for low-P34 trait in soybean accessions PI 567476 and PI 603570A so that molecular markers could enhance breeding efforts to incorporate this important trait into elite soybean cultivars.

Materials and Methods

DNA Sequencing

Genomic DNA was isolated from leaves of 'Williams 82', PI 567476, PI 603570A, and 'Century' using the DNeasy Plant Mini Kit (Qiagen, Inc., Valencia, CA) and used at 5 to 50 ng per polymerase chain reaction (PCR) amplification. Primers were designed to amplify P34 products covering the entire genomic sequence (Jp34G1:CCCCTGCTGGATAATGAAAA and P34R1:AATCCCATGATGCAGGTGGA; p347:AGCAAATCAAAATGGCCAAC and p348:TGGCTTTGCATCTACCCTCT; 34-5:GCACATGCAATAGCAACAGG and 34-6:ACGGCTCAAAGAGGAGAGTG; Jp343pr1:GTCTGCTCGCGTTAAAGGTC and Jp343pr2:TGCTTGACAAATGGAAAGAG; as well as P349:CCCAACCAAAGAGGAATCAG and P3410:TGAAGCATGCATGTTGAAGA). PCR products were analyzed by gel electrophoresis to ensure specific amplification. PCR products were isolated with the Qiaprep Spin Miniprep kit (Qiagen, Inc.) and sequenced with each of the amplification primers at the University of Missouri DNA Core facility.

Molecular Marker Assays and Genotyping

Molecular marker assays were designed to differentiate between wild-type P34 start codon region alleles and the PI 567476/PI 603570A four-basepair insertion alleles. The P34 GC tail assay (Wang et al., 2005) utilized three primers: xp34F1:GCTACAAGTGAAGTGACCATATC; x341:gcgggcACAAGGAAACCCATAACTTGG; and x342:gcgggcagggcgcACAAGGAAACCCATACATAACTTG. Reactions were carried out in 15 μ l; each primer was at 0.375 μ M final concentration in reactions containing template, buffer [40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl₂, 3.75 μ g mL⁻¹ BSA, 200 μ M dNTPs], 5% DMSO, 0.25X SYBR Green I, and 0.2X Titanium *Taq* polymerase (BD Biosciences, Palo Alto, CA). PCR parameters on a DNA Engine Opticon 2 (MJ Research/Bio-Rad, Hercules, California) for the P34 GC Tail assay were as follows: 95 °C for 5 minutes followed by 35 cycles of 95 °C for 20 seconds, 64 °C for 20 seconds, 72 °C for 20 seconds, and then a melting curve from 72 °C to 85 °C.

The fluorescence was read after each cycle and every 0.2 °C with a one second hold during the melt with excitation at 470–505 nm and detection at 523–543 nm. Each genotype produced a product with a characteristic melting profile, as measured by melting temperature (*T_m*) of the negative first derivative of the disappearance of fluorescent signal. Homozygous wild-type P34 samples produced a peak at 75 °C, homozygous mutant P34 alleles produced a peak at 77 °C, and heterozygous P34 alleles produced a peak at 77 °C with a shoulder at 75 °C.

The P34 size assay relied on detecting the four-basepair difference in PCR product sizes for wild-type and mutant P34 alleles. The assay utilized two primers: P34f:CTCACTAATCACTATATATACGACATGC, which was 6-FAM (fluorescein)-labeled on the 5' end; and P34r:ATGGAACGATGAGTTGATATGC. Amplification conditions were 95 °C for 5 minutes followed by 35 cycles of 95 °C for 20 seconds, 60 °C for 20 seconds, 72 °C for 20 seconds. PCR was performed as above except in reaction mix without SYBR Green I Dye, and in 10 μ l reactions with 0.5 μ M each primer. PCR products were diluted 1:50 in water; 1.5 μ l of diluted products were sized on an ABI 3730 DNA analyzer at the University of Missouri DNA Core facility. Wild-type P34 products were 157 bp, mutant products were 161 bp, and heterozygous samples contained both products.

The P34 SimpleProbe assay was based on the dissociation kinetics of an oligonucleotide SimpleProbe (Roche Applied Sciences, Indianapolis, IN) corresponding to the mutant P34 sequence encompassing the four-basepair insertion (Fluorescein-SPC-CACCAAGT-TatgtATGGGTTTCCTTGTGTT-phosphate). The assay utilized the same amplification conditions, the general reaction mixture without SYBR Green I Dye, and the same primers used in the P34 size assay, except P34f was not 6-FAM-labeled. Amplification reactions (20 μ l) consisted of an asymmetric mixture of the amplification primers: 0.2 μ M P34f and 0.5 μ M P34r. The P34 SimpleProbe was included at 0.2 μ M. The dissociation kinetics of the SimpleProbe were assessed following the PCR with the inclusion of a melting curve. For the LightCycler 480 Real-Time PCR System (Roche Applied Sciences), wild-type samples produced a peak at 59 °C, mutant samples produced a peak at 66 °C, and heterozygous samples contained both peaks. For the DNA Engine Opticon 2 (Bio-Rad), wild-type samples produced a broad peak at 58 °C, mutant samples produced a peak at 65 °C, and heterozygous samples contained both peaks.

Plant Materials/Population Development

Two segregating populations were developed from crosses between conventional, P34-containing soybean germplasm and the low-P34 soybean accessions described previously (Joseph et al., 2006): Population 1 (Elite 1 $\times$ PI 567476) and Population 2 (Elite 2 $\times$ PI 603570A). F₂ plants were grown at the Bradford Research and Extension Center located near Columbia, MO in 2007. Parental lines were grown at the same time and

location with the exception of Elite 2, for which Williams 82 was substituted as a wild-type P34 line. Fifty F_2 plants were chosen at random from Population 1 and 100 plants from Population 2; each plant was tagged, and a single leaflet was harvested from each plant and prepared as an FTA card press (Whatman, Clifton, NJ). For Population 1 lines that produced seed, eight lines were genotyped homozygous wild-type, nine were homozygous mutant, and 24 were heterozygous. For Population 2 lines that produced seed, 24 were genotyped homozygous wild-type, 16 were homozygous mutant, and 49 were heterozygous. When plants reached maturity, approximately 20 F_3 seeds were harvested from each tagged F_2 plant in the two populations.

F_3 Seed Genotype and Phenotype Experiment

One F_2 plant from Population 1 that was heterozygous for the mutant P34 allele was threshed at maturity and forty individual F_3 seeds were chipped with a scalpel to provide proteins for Western analysis while allowing the remainder of the seed to be germinated and provide leaf tissue for genotyping.

For protein extraction, each approximately 10-mg seed chip in a 1.5-mL tube was combined with 250 μ L 1X SDS sample buffer (80 mM Tris-Cl, pH 6.8, 2% SDS (v/v), 10% glycerol (v/v), 0.7 M 2-mercaptoethanol, and 0.02 g L⁻¹ bromophenol blue), and the samples were incubated for 30 minutes at room temperature. Plastic pestles were used to carefully macerate seed chips in sample buffer in the tubes. Samples were incubated at 90 °C for ten minutes prior to centrifugation for 5 minutes at 16,000 $\times$ g. Samples were diluted 100-fold by transfer of 5 μ L of supernatant to 495 μ L 1X SDS sample buffer in a fresh tube. Diluted samples were stored at -20 °C. Prior to loading 5 μ L of each sample on a 12.5% acrylamide SDS PAGE gel (Bio-Rad Protean system, Hercules, CA), samples were heated to 90 °C for five minutes and briefly centrifuged. Kaleidoscope Prestained Standards (Bio-Rad) were overlaid in one well per gel. Separated proteins were transferred from the gel to an ImmobilonP transfer membrane (Millipore, Billerica, MA) according to the manufacturer's instructions. Westerns utilizing monoclonal anti-P34 antibodies were performed essentially as described for "Immunoblotting for P34" (Joseph et al., 2006), except the primary antibodies were diluted 1:2500. After the processed membranes dried, the 40 samples on four membranes were subjected to blinded scoring for either high or low intensity P34 bands by three individuals. There was consensus scoring with one exception. That sample was subjected to an independent Western analysis, and was confirmed to be a high intensity P34 sample.

Seed portions containing the embryo were germinated in germination packets (CYG, Mega International, St. Paul, MN). Approximately eight to twelve days after imbibition, one unifoliate leaf from each seedling was pressed onto an FTA card (Whatman, Clifton, NJ). Templates for all genotype PCRs consisted of 1.2 mm washed FTA card punches prepared according to the

manufacturer's instructions. Genotypes were obtained using the GC tail assay.

Genotype Bulk Experiment

Population 1 and Population 2 $F_{2,3}$ lines were assigned a P34 genotype using the P34 GC tail assay. For each population, three F_3 seeds from each line were combined within each genotype class to create a bulk seed sample representing homozygous mutant, homozygous wild-type, and heterozygous P34 genotype classes. Seeds from each genotype class were ground together in a small grinder (SmartGrind, Black & Decker Corp., Towson, MD). In a 1.5 mL tube, 25 mg of each seed sample was combined with 250 μ L of 1X SDS sample buffer, vortexed thoroughly, and heated in a boiling water bath for five minutes prior to centrifugation for five minutes at 16,000 $\times$ g. The supernatants were subsequently diluted 25-fold in 1X SDS sample buffer. Diluted samples were stored at -20 °C. Prior to loading 5 μ L of each sample on each of two 12.5% acrylamide SDS PAGE gels, samples were heated to 90 °C for five minutes and briefly centrifuged. Kaleidoscope Prestained Standards (Bio-Rad) were used in one well per gel. After protein separation, one gel was processed for Western analysis with the P34 antibodies as described above. The duplicate gel was stained with Coomassie Blue R-250 to visualize protein bands.

Results

Identification and Characterization of P34 Gene Sequences from Soybean Cultivars Williams 82 and Century, and Low-P34 Allergen Germplasm Accessions PI 567476 and PI 603570A

The Century (Wilcox, et al., 1980) P34 cDNA sequence (GenBank accession J05560, Kalinski et al., 1990) was initially used in blast searches of the soybean draft genome sequence trace archives in GenBank to identify the soybean genomic DNA region containing the P34 gene. Individual overlapping sequence traces were used to assemble the consensus P34 gene region from Williams 82 (Bernard and Creemens, 1988). The Williams 82 P34 gene consisted of four exons separated by three introns and encompassed 1806 basepairs from start to stop codons (Fig. 1). The Williams 82 P34 genomic region was identical to a genomic DNA sequence annotated without a cultivar description as "*Glycine max* gene for Bd 30K" from GenBank accession AB013289. The Century P34 cDNA sequence deposited as GenBank accession J05560 had several polymorphisms when compared to the exons from the assembled Williams 82 P34 genomic DNA region.

In addition to what appeared to be the authentic P34 gene, two P34 pseudogenes were identified from manual assembly of the soybean trace archive sequences; one P34 pseudogene matched a DNA sequence annotated as "*Glycine max* pseudogene for Bd 30K" from GenBank accession AB013290. Although the DNA sequence identity between the P34 sequences was above 90%, this P34

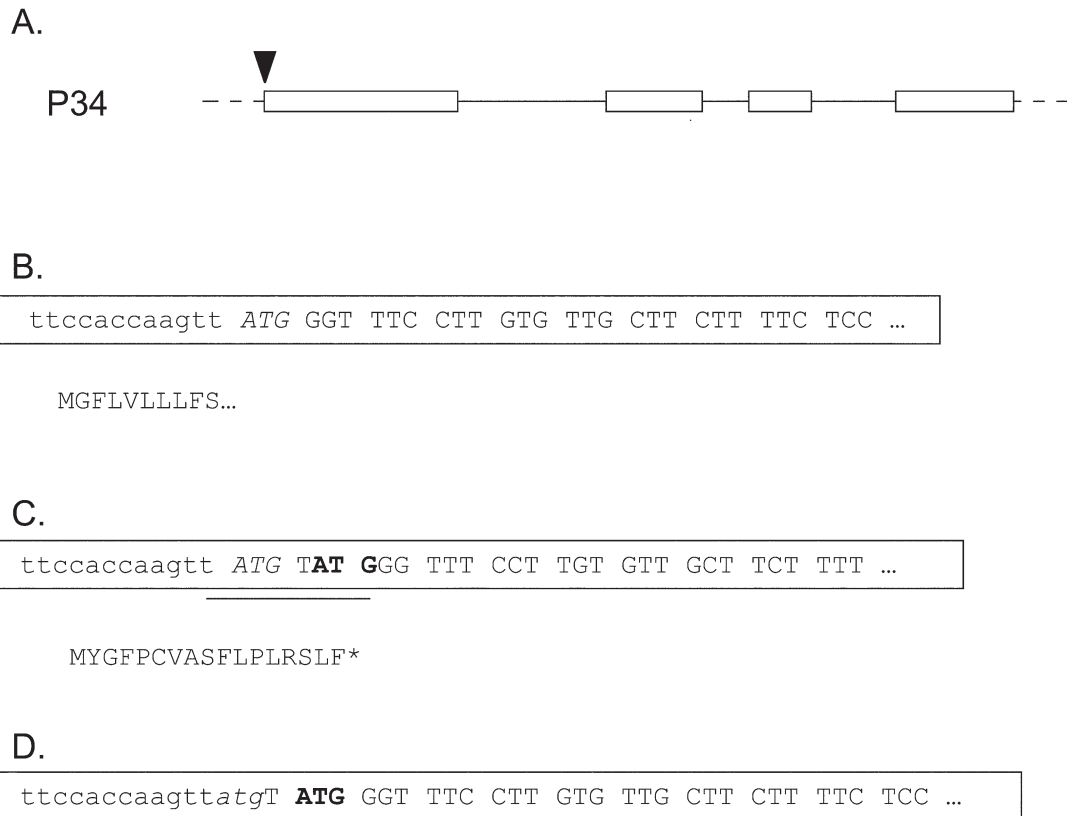


Figure 1. Genomic organization of the P34 gene for Williams 82 and low P34 soybean accessions PI 567476 and PI 603570A. A. Four exons (open boxes) and three introns (lines separating exons) were assembled from trace archive sequence reads and amplification and sequencing of Williams 82 genomic DNA in the P34 gene region. Short sequences of 5' and 3' untranslated regions were also amplified and sequenced as represented by dashed lines. At the start codon of the P34 allele of the low P34 accessions, a four-basepair insertion results in a direct sequence repeat (TATGTATG). The position of the insertion is represented by a solid triangle. B. Wild-type sequence of the P34 gene in the start codon region. 5' untranslated nucleotides are in lowercase letters, the start codon ATG is in italics, and the translated nucleotides are in uppercase letters, with each codon separated by a space. Only the first ten codons are represented. Below the nucleotide sequence is a translation of the first ten amino acids of the wild-type P34 protein. C. Mutant sequence of the P34 allele in the start codon region. One possible effect of the P34 four-basepair insertion allele (the direct repeat is underlined) is translation initiation at the first ATG (italics) and a frameshift producing a 17 amino acid peptide. The second ATG is listed in bold, and the sequence only represents the first ten codons. Below the nucleotide sequence is the frameshifted peptide that would result from translation of the mutant P34 allele. D. Mutant sequence of the P34 allele in the start codon region. Another possible effect of the P34 four-basepair insertion is a reduction or elimination of translation initiation at the second ATG (bold) due to sequence changes just upstream from the start codon, including an additional ATG (italics).

pseudogene contained an inframe stop codon after the first 59 amino acids. Manual analysis of the soybean trace archive sequences and subsequent analysis of the 7X assembly of the soybean genome sequence revealed an additional P34 pseudogene located approximately 10 kilobases away from the authentic P34 gene. This pseudogene contained a stop codon following the first 142 amino acids. The authentic P34 gene (Glycine max 1.01 assembly: Glyma08g12270.1) and one of the pseudogenes (Glyma08g12280.1) were found to reside between microsatellite markers Sat_157 and Sat_212, which corresponded to linkage group A2 (chromosome 08) near the I locus. The pseudogene identical to GenBank accession AB013290 (Glyma05g29104.1) appeared to reside on linkage group A1 (chromosome 05). There was no evidence in the EST collection for expression of either P34 pseudogene.

Using the Williams 82 P34 genomic sequence, PCR primers were designed to amplify the P34 gene region in

overlapping segments from genomic DNA. PCR products corresponding to the P34 region from 267 basepairs upstream from the start codon to 256 basepairs beyond the stop codon were amplified and sequenced from Williams 82, Century, PI 567476, and PI 603570A genomic DNA. Williams 82 and Century P34 sequences were identical to the sequence predicted from trace archives. Alleles for the PI 567476 and PI 603570A P34 gene contained an identical four-basepair insertion at the start codon (Fig. 1). The insertion resulted in a short direct repeat (TATGTATG) that included the original P34 start codon. Genomic sequences for the P34 alleles were deposited in GenBank under accessions FJ616287 for Williams 82 and FJ616288 for PI 567476 and PI 603570A. No other sequence variations were identified among the lines in the P34 gene region.

The four-basepair insertions present in the P34 alleles of PI 567476 and PI 603570A could result in

several possible outcomes: translation initiation from the first ATG codon and a frameshift that would produce a small 17 amino acid peptide, disruption of the translation initiation site due to the change in sequence and the presence of two start codons resulting in no or reduced translation, or unaltered translation of the P34 gene (Fig. 1). Thus, the PI 567476 and PI 603570A P34 alleles can be considered mutant alleles when compared to the wild-type P34 gene allele present in Williams 82.

Development of Molecular Marker Assays for PI 567476 and PI 603570A P34 Alleles

Access to genomic DNA sequence of the two P34 pseudogenes, the authentic P34 gene, and the PI 567476 and PI 603570A alleles of the P34 gene enabled us to design

and validate three different molecular marker assays for analysis of the P34 genotype: the P34 GC tail assay, the P34 size assay, and the P34 SimpleProbe assay (Fig. 2). Each of these three molecular marker assays was capable of efficient genotyping of the P34 allele present in plant samples, but since there is no standard genotyping platform in place, we evaluated different assay types.

One assay (P34 GC tail assay) was based on allele-specific PCR amplification in the presence of the dye Sybr Green I of the wild-type or mutant alleles of P34. This assay reliably distinguished plants containing mutant P34 alleles from those that were homozygous for the wild-type P34 alleles and those that were heterozygous for the P34 alleles, although the results were best with highly purified DNA templates (Fig. 2B). The assay

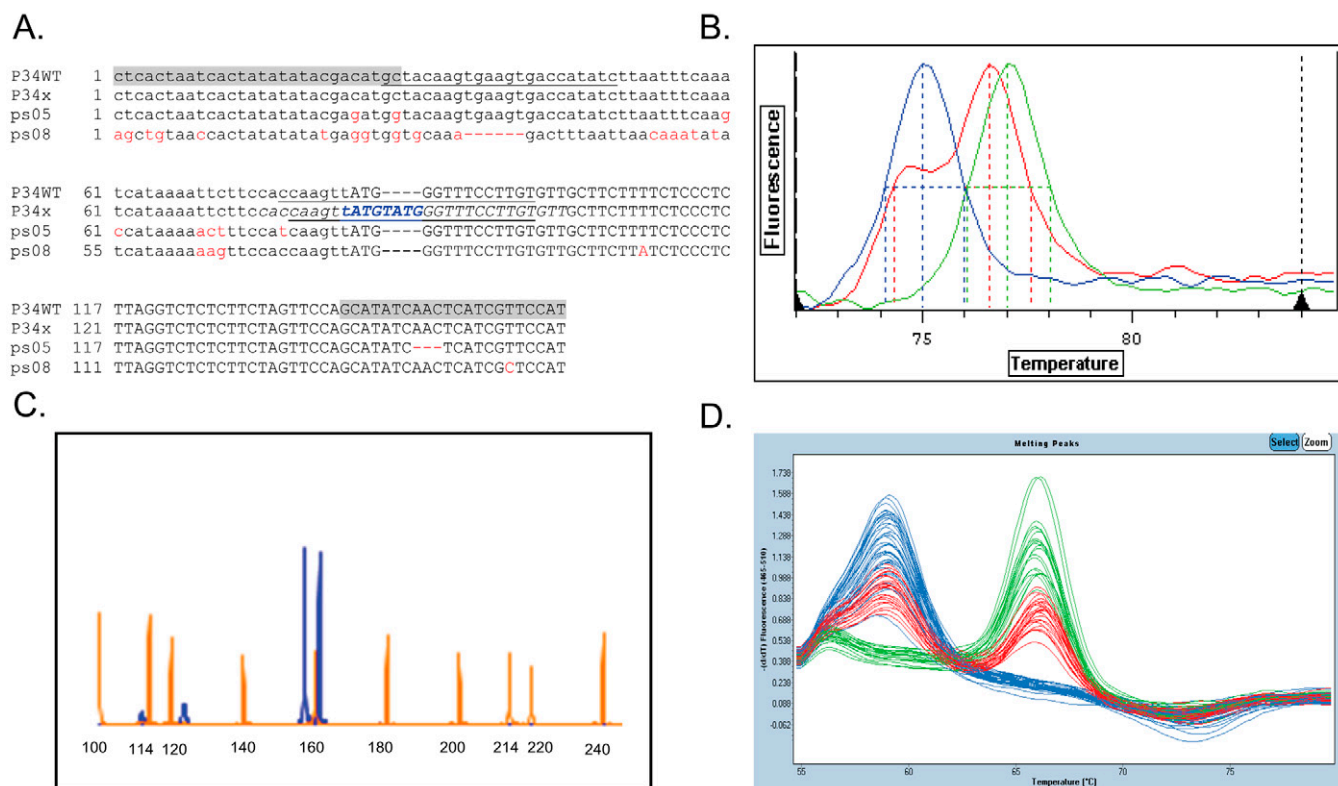


Figure 2. Molecular marker assays for wild-type or mutant P34 alleles. A. Sequence alignment encompassing all of the primers utilized in the P34 molecular marker assays. P34WT represents Williams 82 genomic DNA sequence on soybean chromosome 8; P34x represents PI 567476 and PI 603570A mutant P34 alleles; ps05 and ps08 represent the P34 pseudogenes on soybean chromosomes 5 and 8, respectively. Lower case text represents the 5'-untranslated region, and the first exon sequences are in uppercase text. Red sequences indicate bases that differ for one or both of the P34 pseudogenes. Blue bolded bases represent the direct repeat resulting from the four-basepair insertion in PI 567476 and PI 603570A alleles. Gray highlighted sequences represent the regions corresponding to the primers used for the P34 size assay and the SimpleProbe assay. The italicized bases in the P34x sequence represent the SimpleProbe sequence. Underlined sequences in the P34WT and P34x sequences represent the regions corresponding to the primers used for the GC tail assay. These assay primers also contained tails (described in Materials and Methods). B. Example of genotype output from the P34 GCtail assay. Melting curves generated from the negative first derivative SYBR Green I fluorescence disappearance over temperature. The blue peak is from a wild-type P34 sample, the red peaks are from a heterozygous P34 sample, and the green peak is from a four-basepair insertion P34 sample. C. Example of genotype output from the P34 size assay. Shown is fragment analysis output of GeneScan 600 LIZ standards (orange) with their sizes listed on the x-axis. The blue peaks represent a heterozygous P34 sample which produced two intense peaks on either side of the 160 bp standard, with the 157 bp product representing the wild-type P34 allele product, and the 161 bp product representing the four-basepair insertion P34 product. D. Example of genotype output from the P34 SimpleProbe assay. Melting curves generated from the negative first derivative of SimpleProbe fluorescence disappearance over temperature. The blue peaks are from wild-type P34 samples, the red peaks are from heterozygous P34 samples, and the green peaks are from four-basepair insertion P34 samples.

requires access to a real-time PCR instrument but has very low reagent costs.

A second molecular marker assay was developed that took advantage of the insertion of four basepairs in the P34 gene in the mutant allele (P34 size assay). P34 PCR amplification products were diluted and subjected to fragment analysis to determine the exact size of PCR products, which differed by four basepairs depending on the P34 alleles that were present (Fig. 2C). This assay was robust from different types of DNA templates but required post-PCR processing and a fragment analysis system to identify the four basepair difference in amplicon sizes.

The third molecular marker assay was based on melting curve analysis of a Roche SimpleProbe designed to the mutant P34 allele (SimpleProbe assay). Each genotype class produced a characteristic melting curve profile that was distinguishable on the Roche 480 LightCycler and a standard real-time PCR instrument (Fig. 2D). Similar to the P34 GC tail assay, the SimpleProbe assay required access to a real-time PCR instrument, but it has slightly higher reagent costs.

The P34 GC tail assay was used to investigate the occurrence of the mutant P34 allele in a subset of soybean lines that were major contributors to the North American elite soybean germplasm pool (Gizlice et al., 1994; Sneller, 1994). None of the seventeen “ancestral” soybean lines contained the PI 567476/PI 603570A mutant P34 allele (data not shown).

Association Analysis of F_3 Genotypes and F_3 Seed Phenotypes in a Population Segregating for the Low-Allergen Trait

The mutant allele of the P34 gene in the two identified low-P34 lines became the primary candidate for the molecular genetic basis of the low-P34 phenotype. To investigate the association of the P34 genotype with the P34 protein phenotype, segregating populations were developed from crosses between elite lines containing wild-type levels of the P34 protein and the low-P34 lines PI 567476 and PI 603570A. For Population 1 (Elite 1 $\times$ PI 567476) an F_2 plant was identified by genotype that was heterozygous for the P34 allele. The F_3 seeds from this heterozygous F_2 plant were harvested at maturity. From a subset of the seeds, a small portion of each seed was removed with a scalpel (chipped) to provide proteins for Western analysis of the P34 protein; the remainder of each seed was germinated and genotyped with the P34 GC tail assay for either the wild-type, four-basepair insertion P34 allele, or both. For the 40 samples analyzed, the ratio of homozygous wild-type:heterozygous:homozygous mutant P34 genotypes was 7:23:10.

Phenotypes were determined using P34 monoclonal antibodies and Western blotting of SDS-PAGE-separated total-extracted seed chip proteins. Developed Western blots were subjected to blinded scoring for either high- or low-intensity P34 cross-reacting protein bands. All lines scored to contain low-P34 protein were homozygous for the mutant P34 allele genotype, reflecting a perfect association

between the P34 mutant allele genotype and the low-P34 phenotype. Lines which scored high P34 protein contained either the wild-type or heterozygous P34 genotype.

Association Analysis of Genotype Bulks

Because of variable protein extraction efficiency, seed chips were not the optimum tissue to determine the phenotype for the P34 trait. Therefore, for Population 1 and a second population (Elite 2 $\times$ PI 603570A, Population 2), we genotyped F_2 plants and harvested their F_3 seeds for phenotypic analysis. Lines were categorized based on their F_2 P34 genotype class (homozygous mutant, homozygous wild-type, or heterozygous). Within each of the three genotype classes, three F_3 seeds representing each F_2 plant were combined to make bulks. Proteins were extracted from ground seed samples from each of the classes and analyzed by Western blotting for the P34 protein. The parental lines were included in the Westerns for comparisons. Parental line Elite 2 was not available, so Williams 82 was used as a substitute for the wild-type parent for Population 2. In both Populations, results demonstrated that F_2 plants which were homozygous for mutant P34 alleles produced seeds with the low-P34 phenotype, similar to their low-P34 parent (Fig. 3). With the bulked samples there was also a perfect association between the P34 mutant allele genotype and the low-P34 phenotype. Observed differences in P34 protein accumulation between $F_{2,3}$ lines that were heterozygous and homozygous for the P34 wild-type allele were not explored further.

Evaluation of Reduction in Seed P34 Protein Levels

We routinely utilized enough seed-protein extract in our experiments to easily detect the P34 protein band in mutant parental line samples, although at an obviously less intense level than in wild-type samples. When we evaluated the difference in P34 levels in Westerns after dilution of the wild-type protein samples, the P34 band intensities detected from protein extracts of ground whole-seed samples were similar when the parental wild-type protein sample was diluted eightfold compared to the low-P34 parent PI 603570A (Fig. 4). Apparent differences between segregating mutant bulk sample and segregating wild-type samples were also approximately eightfold when evaluated with dilutions of the wild-type protein sample. Similar differences were observed for the PI 567476 samples compared to wild-type (data not shown).

DISCUSSION

Soybean is an important source of high protein meal that is incorporated into many foods and feeds. Soybean seeds contain multiple proteins that are considered to be allergenic to humans. In addition, livestock such as weanling pigs (*Sus scrofa domestica*) have been shown to have sensitivity to soybean meal proteins (Li et al., 1991; Li et al., 1990). The dominant soybean allergen is Gly m Bd 30K, a cysteine protease-type protein also known as

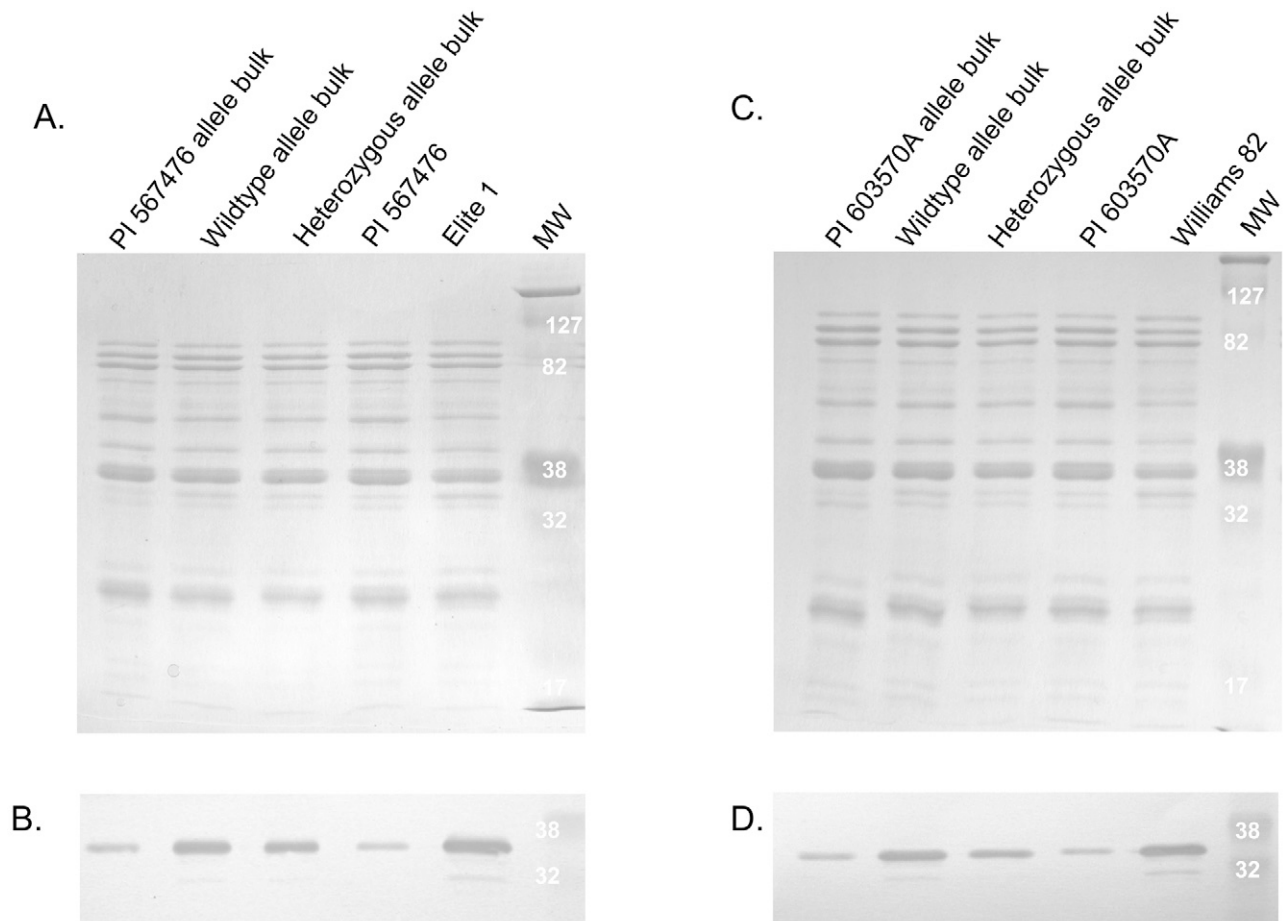


Figure 3. Association of the low P34 phenotype with the PI 567476 and PI 603570A P34 four-basepair insertion alleles. A. Coomassie Blue stained total seed proteins extracted from ground bulk seed samples and separated by SDS-PAGE from Population 1 and the parental lines PI 567476 and Elite 1. The allele bulks consisted of mixed ground seed from the indicated genotype class listed and each lane represents the proteins extracted from 50 μ g of dried seed corrected for dilution. MW indicates protein molecular weight standards, and the individual proteins are labeled with their apparent molecular mass. B. Western blot with anti-P34 monoclonal antibodies on duplicate gel, as in panel A. C. Coomassie Blue stained total seed proteins extracted from ground bulk seed samples and separated by SDS-PAGE from Population 2 and the parental line PI 603570A as well as the wild-type P34 cultivar Williams 82. Each lane is labeled similar to those in panel A. D. Western blot with anti-P34 monoclonal antibodies on duplicate gel, as in panel C.

P34 (Kalinski et al., 1990; Ogawa et al., 1993). Although the P34 protein is not particularly abundant in soybean seeds, its presence is almost uniformly conserved in diverse germplasm (Joseph et al., 2006; Xu et al., 2007; Yaklich et al., 1999). Suppression of P34 protein expression in transgenic soybeans demonstrated no negative effects from elimination of the P34 protein in seed development and final seed composition (Herman et al., 2003). Large scale screening of the USDA soybean germplasm collection led to identification of two *G. max* lines with reduced P34 content (Joseph et al., 2006). While these two soybean accessions were shown to accumulate greatly reduced levels of the P34 protein in mature seeds, an understanding of the molecular genetic basis for this trait was lacking.

Results presented here demonstrated that the low-P34 accessions PI 567476 and PI 603570A each contain an identical four-basepair insertion at the P34 start codon. No other sequence differences were observed in the P34 genomic DNA region between the two low-P34 germplasm accessions and the genomic sequence of the

standard cultivar Williams 82. When the P34 cDNA was amplified and sequenced from the two low-P34 germplasm accessions (Joseph et al., 2006), the forward amplification primer initiated two bases upstream of the original start codon, and thus the resulting product would not have included the four-basepair insertion. We have not been able to detect the P34 allele containing the four-basepair insertion from a subset of ancestral soybean lines (Sneller, 1994) or from any other lines unrelated to the two low-allergen accessions.

The lack of any reduction in P34 mRNA levels for the two mutant lines (Joseph et al., 2006) would be consistent with a model in which the P34 protein is mistranslated due to a frameshift, or in which translation initiation is reduced due to alteration of sequence at the start codon. The exact molecular mechanism remains elusive; however the accumulation of reduced levels of full-length P34 protein implicates reduced translation due to the insertion as the likely cause. Although we identified two P34 pseudogenes, neither one appeared capable of encoding

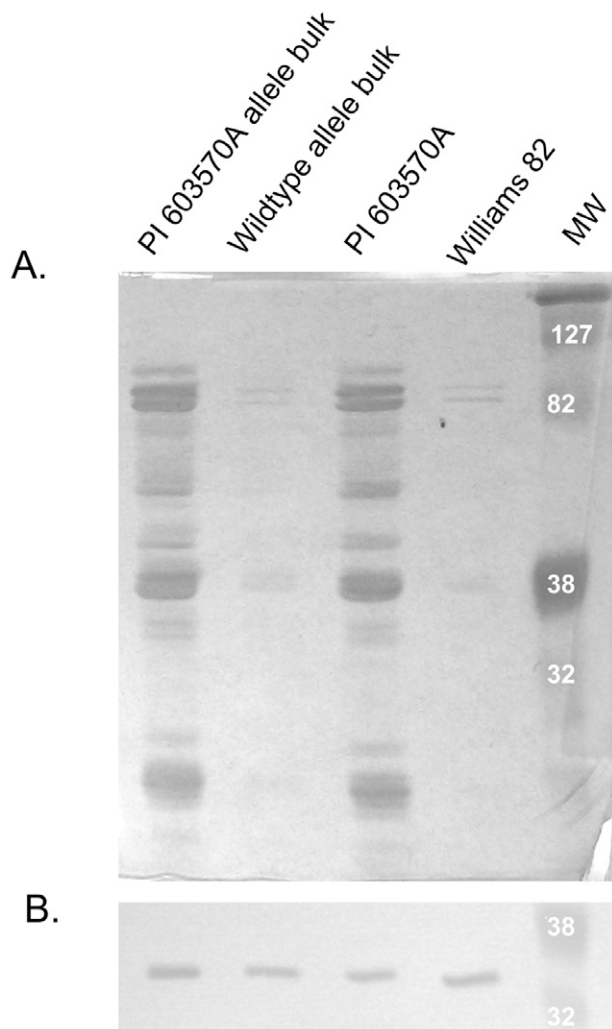


Figure 4. Reduction of P34 protein level in four-basepair-insertion P34 lines. **A.** Coomassie Blue stained total seed proteins extracted from ground bulk seed samples representing 50 μ g dry ground seed and separated by SDS-PAGE from the PI 603570A and the mutant P34 allele bulk sample from Population 2. The cultivar Williams 82 and the wild-type P34 allele bulk samples were diluted an additional eightfold (representing 6.25 μ g dried seed). The allele bulks consisted of mixed ground seed from the indicated genotype class. MW indicates protein molecular weight standards, and the individual proteins are labeled with their apparent molecular mass. **B.** Western blot with anti-P34 monoclonal antibodies on duplicate gel, as in panel A.

a full length protein that could account for the P34 band that appeared in Westerns. Indeed, in our experiments, there was an approximately eightfold reduction in P34 protein accumulation analyzed by Westerns for the two germplasm accessions and derived lines homozygous for the mutant P34 allele compared to standard cultivars. We observed a perfect association of the inheritance of the mutant P34 alleles with the low-P34 allergen phenotype. The four-basepair insertion does not appear to condition a complete null allele of the P34 gene.

The PI 567476 and PI 603570A P34 alleles were completely associated with the low-P34 seed phenotype in independent segregating populations. Molecular marker

assays were developed that detected the P34 genotype based on the presence or absence of the four-basepair insertion at the start codon. Molecular marker assays were capable of distinguishing homozygous mutant, wild-type, and heterozygous plants. Although the heterozygous seeds seemed to produce an intermediate P34 phenotype, the technical aspects of accurately phenotyping individual heterozygous seeds would be very challenging. Use of the P34 molecular markers for direct selection of the mutant P34 allele would allow the most rapid incorporation of the low-P34 trait into elite germplasm through use of backcross breeding and identification and selection of individuals with heterozygous P34 alleles. The low-allergen phenotype could be recovered at the desired generation after selection of segregating individuals containing homozygous mutant P34 alleles.

Acknowledgments

Excellent technical support was provided by Paul Little, and soybean field and breeding management was performed by Kerry M. Clark.

References

- Bernard, R.L., and C.R. Creemans. 1988. Registration of Williams 82 soybean. *Crop Sci.* 28:1027–1028.
- Gizlice, Z., T.E. Carter, and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Sci.* 34:1143–1151.
- Herman, E.M., R.M. Helm, R. Jung, and A.J. Kinney. 2003. Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol.* 132:36–43.
- Joseph, L.M., T. Hymowitz, M.A. Schmidt, and E.M. Herman. 2006. Evaluation of Glycine germplasm for nulls of the immunodominant allergen P34/Gly m Bd 30k. *Crop Sci.* 46:1755–1763.
- Kalinski, A., J.M. Weisemann, B.F. Matthews, and E.M. Herman. 1990. Molecular cloning of a protein associated with soybean seed oil bodies that is similar to thiol proteases of the papain family. *J. Biol. Chem.* 265:13843–13848.
- Li, D.F., J.L. Nelssen, P.G. Reddy, F. Blecha, R. Klemm, and R.D. Goodband. 1991. Interrelationship between hypersensitivity to soybean proteins and growth performance in early-weaned pigs. *J. Anim. Sci.* 69:4062–4069.
- Li, D.F., J.L. Nelssen, P.G. Reddy, F. Blecha, J.D. Hancock, G.L. Allee, R.D. Goodband, and R.D. Klemm. 1990. Transient hypersensitivity to soybean meal in the early-weaned pig. *J. Anim. Sci.* 68:1790–1799.
- Ogawa, T., N. Bando, H. Tsuji, H. Okajima, K. Nishikawa, and K. Sasaoka. 1991. 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. (Tokyo)* 37:555–565.
- Ogawa, T., H. Tsuji, N. Bando, K. Kitamura, Y. Zhu, H. Hirano, and K. Nishikawa. 1993. Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. *Biosci. Biotechnol. Biochem.* 57:1030–1033.
- Sneller, C.H. 1994. Pedigree analysis of elite soybean lines. *Crop Sci.* 34:1515–1522.
- Wang, J., K. Chuang, M. Ahluwalia, S. Patel, N. Umblas, D. Mirel, R. Higuchi, and S. Germer. 2005. High-throughput SNP genotyping by single-tube PCR with Tm-shift primers. *Biotechniques* 39:885–893.
- Xu, C., T.J. Caperna, W.M. Garrett, P. Cregan, H. Bae, D.L. Luthria, and S. Natarajan. 2007. Proteomic analysis of the distribution of the major seed allergens in wild, landrace, ancestral, and modern soybean genotypes. *J. Sci. Food Agric.* 87:2511–2518.
- Yaklich, R.W., R.M. Helm, G. Cockrell, and E.M. Herman. 1999. Analysis of the distribution of the major soybean seed allergens in a core collection of *Glycine max* accessions. *Crop Sci.* 39:1444–1447.
- Wilcox, J.R., K.L. Athow, T.S. Abney, F.A. Laviolette, and T.E. Richards. 1980. Registration of 'Century' soybean. *Crop Sci.* 20:415.