The Low Phytic Acid Phenotype in Soybean Line CX1834 Is Due to Mutations in Two Homologs of the Maize Low Phytic Acid Gene

Jason D. Gillman, Vincent R. Pantalone, and Kristin Bilyeu*

Abstract

Plant seeds accumulate phosphorus in the form of myo-inositol-1,2,3,4,5,6-hexakisphosphate, commonly referred to as phytic acid. Phytic acid is found complexed with cationic mineral species in the form of phytate, which is not well digested or absorbed by monogastric species such as humans, poultry, and swine. As a result, soybean [Glycine max (L.) Merr.] has an effective deficiency of phosphorus and other minerals, despite high levels of minerals and phosphorus in the seed. Excreted phytate can also contribute to phosphorus contamination of groundwater and eutrophication of freshwater lakes and streams. In maize [Zea mays ssp. mays], a recessive mutation in a conserved region within the low phytic acid 1 (lpa1) gene is responsible for the low phytic acid phenotype. We have identified recessive mutations in two soybean homologs of the maize lpa1 gene in soybean line CX1834, a mutagenized line with a low phytic acid phenotype. In three populations analyzed, we identified complete association between homozygosity for mutant alleles of the two lpa1 homologs and the low phytic acid phenotype in soybean. Molecular marker assays were designed that can be used to directly select for the mutant alleles that control the phenotype.

J.D. Gillman and K. Bilyeu, USDA-ARS, Univ. of Missouri-Columbia, 110 Waters Hall, Columbia, MO 65211; V.R. Pantalone, Dep. of Plant Sciences, 2431 Joe Johnson Dr., Univ. of Tennessee, Knoxville, TN 37996-4561. Mention of a trademark, vendor, or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products or vendors that may also be suitable. Received 21 Jan. 2009. *Corresponding author (bilyeuk@missouri.edu).

Abbreviations: ABC, ATP-binding cassette; chr, chromosome; LG, linkage group; lpa1, low phytic acid 1; MIPS, myo-inositol phosphate synthase; MRP, multidrug resistance protein; NIL, near-isogenic line; PCR, polymerase chain reaction; QTL, quantitative trait loci; RIL, recombinant inbred line; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; W82, ‘Williams 82’.
Seed phytic acid content and inorganic phosphate content have been shown to have an inverse relationship to each other (Bilyeu et al., 2008; Chen et al., 1956; Gao et al., 2008; Raboy et al., 2000; Shi et al., 2007; Wilcox et al., 2000). The development of rapid, simple assays to evaluate whole or partial seeds for inorganic phosphate has led to the development and identification of mutants in soybean (Hitz et al., 2002; Wilcox et al., 2000), maize (Zea mays L. ssp. mays) (Pilu et al., 2003; Raboy et al., 2000), and barley (Hordeum vulgare L.) (Larson et al., 1998) with lowered phytic acid.

Mutagenesis by Wilcox et al. (2000) with the soybean breeding line CX1515-4 yielded two independent mutants: M153 and M766. Upon testing of F1 seeds, both lines displayed elevated inorganic phosphate, and a concomitant decrease in phytic acid. These results were very similar to those previously demonstrated for the barley (Larson et al., 1998) and maize low phytic acid 1 (lpa1)-1 (Raboy et al., 2000) mutants. Because the M153 line was shown to have a greater reduction in phytic acid, M153 was used in the development of several independent breeding lines, including the CX1834 series (CX1834-1-2, CX1834-1-3, and CX1834-1-6; Otłmans et al., 2004; Walker et al., 2006; Wilcox et al., 2000).

Although the initial evaluation of low phytic acid soybean line M153 by Wilcox et al. (2000) suggested that only a single locus was responsible for the low phytic acid phenotype, several genetic studies have implicated two independent recessive loci responsible for the low phytic acid phenotype (Gao et al., 2008; Otłmans et al., 2004; Walker et al., 2006). Because the myo-inositol phosphate synthase (MIPS) gene MIPS1 had been shown to be responsible for the low phytic acid trait in an independent soybean mutant line (Hitz et al., 2002), the sequences of all four soybean MIPS genes were characterized for the low phytic acid line CX1834. No mutations were identified in the MIPS genes, and the MIPS gene did not associate with either the low phytic acid phenotype nor the two loci linked to the trait (Chappell et al., 2006); subsequent mapping studies confirmed the exclusion of the MIPS genes as candidates for the low phytic acid trait in CX1834 (Maroof et al., 2009). Genetic mapping studies revealed two quantitative trait loci (QTL) for the low phytic acid trait from CX1834, one on linkage group (LG) N and one on LG L, with the LG N locus either closer to the QTL marker identified or resulting in a larger impact on seed phytic acid (Gao et al., 2008; Maroof et al., 2009; Scaboo et al., 2009; Walker et al., 2006). The genetic complexity of the trait was revealed by the absolute requirement for two contributing loci containing independent but interacting recessive alleles for expression of the low phytic acid/high inorganic phosphate phenotype (Otłmans et al., 2004).

Maize low-phytate lpa1 mutants were subjected to a fine-mapping study that identified a multidrug resistance–associated protein (MRP) ATP-binding cassette (ABC) transporter-encoding gene as causative for the low phytic acid phenotype. Three independent low phytic acid Mu-insertion lines contained transposon insertion gene-disruption events, and the maize lpa1-1 allele contained a missense mutation in the same transporter gene. The authors name this MRP ABC transporter-encoding gene MRP4, and the mutant form is referred to as low phytic acid 1 (Shi et al., 2007).

The maize lpa1 gene is most closely related to the MRP5 gene in Arabidopsis (Shi et al., 2007). The MRP5 is a gene of unknown function, but it contains an integral transmembrane domain and a cytosolic ATP-binding domain (Klein et al., 2006). Gene silencing by suppression of the maize MRP4 gene, lpa1, in transgenic maize resulted in a low-phytate phenotype (Shi et al., 2007). The authors also identified a soybean homolog of the maize lpa1 gene and created transgenic soybeans with gene silencing of the lpa1 homolog. As in the situation with maize, transgenic soybeans with gene silencing of the lpa1 homolog produced seeds containing a low phytic acid and increased seed inorganic phosphate phenotype (Shi et al., 2007).

The objective of this work was to use the soybean homologs of the maize lpa1 gene as candidate genes for the underlying basis of the low phytic acid/high inorganic phosphate phenotype in the soybean lines CX1834 and M766. With the availability of a new whole-genome shotgun soybean genome assembly (http://www.phytozome.net/soybase [verified 4 May 2009]), we were able to identify two soybean homologs of the maize lpa1 gene and Arabidopsis MRP5 gene: Glyma03 g32500 and Glyma19 g35230. We discerned that these genes are in close proximity (www.soybase.org [verified 4 May 2009]) to the markers previously identified as linked to low phytic acid loci in CX1834 by QTL analysis (Gao et al., 2008; Scaboo et al., 2009; Walker et al., 2006). Following the same general strategy, another group recently independently identified one of the soybean lpa1 homologs as a potential candidate for the low phytic acid trait in CX1834 (Maroof et al., 2009). Their major finding was the identification of a nonsense mutation in an lpa1 homolog in close proximity to the LG N/chromosome (chr) 3 QTL region identified in CX1834. Although a soybean lpa1 homolog was identified on LG L/chr 19, characterization of the CX1834 allele of the gene was not reported; analysis of only one of the two genes responsible for the phenotype prohibited definitive conclusions (Maroof et al., 2009). We independently examined the complete genomic sequence of the lpa1 homologs residing near the LG N and LG L low phytic acid QTL from the soybean low phytic acid line CX1834 and identified the nonsense mutation in the CX1834 allele of the LG N/chr 3 lpa1 homolog. In addition, we have identified a missense mutation in a conserved amino acid in the CX1834 allele of the LG L/chr 19 lpa1 homolog. High-throughput molecular marker assays were designed for each of the two mutations. Our discovery of both mutations and the subsequent development of mutation-specific molecular marker assays allowed us to perform comprehensive association analyses: a perfect association was identified.
between the two recessive mutations and the low phytic acid/high inorganic phosphate phenotype in three independent populations segregating for the low phytic acid trait. During analysis of the segregating populations, we observed several independent recombinant progeny wherein the linkage between the \( Lpa1 \) locus and a tightly linked (<0.4 cM) simple sequence repeat (SSR) marker was broken. In addition, we report the discovery of two novel alleles of the \( lpa1 \) homologs present in the alternate low phytic acid soybean line M766.

**Materials and Methods**

**DNA Isolation and PCR for Sequencing of \( Lpa1 \) and \( Lpa2 \)**

Genomic DNA was isolated from ~30 mg of leaf or dried seed tissue using the DNeasy Plant Mini Kit (Qiagen, Inc., Valencia, CA) and used at 5 to 50 ng per polymerase chain reaction (PCR) amplification. Polymerase chain reaction products were analyzed by gel electrophoresis to ensure specific amplification. Polymerase chain reaction was performed using Ex Taq according to manufacturer's recommendations (Takara, Otsu, Shiga, Japan) in a PTC-200 thermocycler (MJ Research/Bio-Rad, Hercules, CA) using the following conditions: 95°C for 5 min followed by 40 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 4 min. Polymerase chain reaction products were verified for size by running on a 1% agarose gel. Polymerase chain reaction products were isolated with the Qiaprep Spin Miniprep kit (Qiagen, Inc.) and sequenced at the University of Missouri DNA Core facility. The genomic region for \( Lpa1 \) and \( Lpa2 \) was divided into three regions for amplification. Primers used for amplification and sequencing are detailed in Supplementary Table 1.

**Sequence Analysis**

Sequencing traces were imported into ContigExpress (Invitrogen, Carlsbad, CA), trimmed, aligned, and manually evaluated for disagreements between "Williams 82" (W82) reference (http://www.phytozome.net/soybean) and sequencing trace contigs. Putative single nucleotide polymorphisms (SNPs) were verified by at least two independent PCR reactions. Sequences were aligned using the AlignX software (Invitrogen).

**Development of the \( F_{58} \) RIL Population**

A set of 144 randomly selected \( F_{58} \) recombinant inbred lines (RILs) from a cross of 560IT \( \times \) CX1834 (Walker et al., 2006) were examined for association between free phosphate levels and the presence of the CX1834 alleles of \( Lpa1 \) and \( Lpa2 \). 560IT is an \( F_6 \) derived Maturity Group V cultivar from a cross of Hutcheson \( \times \) TN89-39 (Pantalone et al., 2003), and CX1834 is a low phytic acid, Maturity Group III germplasm developed by J.R. Wilcox of the USDA-ARS at Purdue University. DNA isolation, genotyping for chr3/LG N (Satt237) and chr19/LG L (Satt561) SSR markers and quantitation of free phosphate for each RIL was performed as previously described (Walker et al., 2006).

**Development and Inorganic Phosphate Quantification of the \( F_{9,10} \) NIL Population**

The second population we utilized for association tests was derived from one of the \( F_{58} \) RILs (Walker et al., 2006), designated 56Cx-526, which was derived from an \( F_6 \) plant found to be a double heterozygote at the two high-phosphate linked SSR loci (Scaboo et al., 2009). Seed was harvested in bulk and carried forward for this experiment. In December 2004, 180 seeds of that line were sown in 60 hills at 3 seeds hill\(^{-1}\) numbered as TN05PR-286 to TN05PR-345 at the USDA Tropical Agriculture Research Station in Isabela, PR. Leaf tissue samples from each individual \( F_{9,10} \) plant were pressed to store plant DNA on FTA cards (Whatman, Inc., Florham Park, NJ) and kept for later genetic analysis. Each of the three plants hill\(^{-1}\) was harvested individually and designated as an \( F_{9,10} \) near-isogenic line (NIL) TN05PR-286-1, TN05PR-286-2, TN05PR-286-3, TN05PR-287-1, TN05PR-287-2, TN05PR-287-3, etc.

The \( F_{9,10} \) NIL population was examined for genotype at SSR markers Satt237 and Satt561 as previously described (Walker et al., 2006), using stored plant presses on FTA cards as DNA source, according to manufacturer’s recommendations (Whatman, Inc.). These SSR markers have been shown to be linked to the low phytic acid loci on linkage group N and L, respectively. Seed from a subset of this population, representing 7 to 10 plants for each of all possible combinations of homozygous genotypes for Satt237 and Satt561 were examined for inorganic phosphate levels and \( Lpa1 \) and \( Lpa2 \) genotypes. Three or four individual seed plants were characterized for inorganic phosphate phenotype. Approximately one-third of the seed was excised with a clean razor and each seed chip was individually evaluated for seed inorganic phosphate levels as previously described. The remainder of the seed was ground to a fine powder in a mortar and pestle with liquid nitrogen, and subjected to individual seed DNA extraction (Qiagen DNeasy Plant Kit, Carlsbad, CA) for \( Lpa1 \) and \( Lpa2 \) genotyping assays. As a screen for potential recombination events between Satt237 and \( Lpa1 \), four seeds from 51 plants homozygous for the 560IT allele of Satt237 and homozygous for the CX1834 allele of Satt561 were examined for inorganic phosphate phenotype. The two potentially recombinant plants that displayed unexpected high phosphate levels for their four progeny seed were evaluated for \( Lpa1 \) and \( Lpa2 \) genotypes.

Levels of inorganic phosphate were quantified by slight modifications of the method described by Wilcox et al. (2000). Single seeds were either ground to a powder in liquid nitrogen or chipped with a scalpel to remove an ~10- to 30-mg portion distal to the axis. As a further control, single-seed chip analysis was compared with that of the remaining seed that was processed by grinding. No differences in phosphate values were detected (Bilyeu
et al., 2008). Ground samples (10–30 mg) or intact seed chips were extracted overnight by shaking at 4°C in 0.5 mL of extraction buffer (12.5% [w/v] trichloroacetic acid, 25 mM MgCl₂). Particulates were allowed to settle at room temperature for 30 min to 1 h. Ten microliters of sample supernatant was combined with 90 μL of water and 100 μL of colorimetric reagent (1 volume of 3 M H₂SO₄, 1 volume of 0.02 M ammonium molybdate tetrahydrate, 1 volume of 10% [v/v] ascorbic acid, and 2 volumes of water) in a 96-well plate well, allowed to incubate at room temperature 1.5 h, and then analyzed at 825 nm in a Versamax tunable microplate reader (Molecular Devices, Sunnyvale, CA). A standard curve was generated with K₂HPO₄ standards. Results were converted to μg P g⁻¹ seed.

Confirmation of Satt237 Marker Phenotype for a Subset of F₉₁₀ NIL Population

Despite the homozygous genotype designation of the samples for both high-phosphate linked SSR markers, some samples previously genotyped as homozygous for the CX1834 allele of Satt237 were noted to be segregating for Lpa1 (and inorganic phosphate). For those samples for which this occurred, we retested the Satt237 marker genotype. This was performed as previously described (Walker et al., 2006), with slight modification: DNA was extracted from ~30 mg pulverized seed tissue with Qiagen DNeasy Plant Kits and PCR was performed in a PTC-200 thermocycler (MJ Research/Bio-Rad) using the following conditions: 95°C for 5 min followed by 40 cycles of 95°C for 30 s, 47°C for 30 s, and 72°C for 30 s. Amplification primers for Satt237 were as described previously (Song et al., 2004) (http://bldg6.arsusda.gov/~pooley/soy/cregan/soy_map1.html [verified 4 May 2009]). Reactions were performed in 20 μL containing 5 to 50 ng DNA template, primers, buffer (40 mM Tricine-KOH [pH 8.0], 16 mM KCl, 3.5 mM MgCl₂, 3.75 μg mL⁻¹ BSA, 200 μM dNTPs), 5% DMSO, and 0.2× Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). Amplicons for the Satt237 marker (Song et al., 2004) were submitted to the DNA Core Facility of the University of Missouri (Columbia, MO) and separated on an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA). Results were interpreted using the Peak Scanner software v1.0 (Applied Biosystems).

CX1834 × S97-1688 Population Development

The third population was developed by crossing CX1834-1-2 with a competitive-yielding Maturity Group V line, S97-1688. The crosses were made in Knoxville, TN, and the F₁ seeds were sent to Costa Rica for generation advancement to the F₂ stage via single-seed descent (Fehr, 1991). A total of 186 RILs were randomly chosen from the S97-1688 × CX1834-1-2 population. One of the F₅₉ lines from that population was designated as TN05-5109. At harvest, 94 F₁₂₃ lines derived from TN05-5109 were threshed as single plants, and the seeds from each line were qualitatively screened for inorganic phosphate phenotype. Six lines were identified to have a high-phosphate phenotype, and four individual seeds from each of those lines were chipped for quantitative phosphate analysis. The remainder of each seed was germinated for genotyping. Analysis of inorganic phosphate content was performed exactly as described previously for the F₉₁₀ NIL population.

SimpleProbe Genotyping Assays

The SimpleProbe assays are based on the disassociation kinetics of SimpleProbe oligonucleotides (Roche Applied Sciences, Indianapolis, IN) designed using the Lightcycler Probe Design Software, version 1 (Roche Applied Sciences) to be exactly complementary to the W82 reference sequence (http://www.phytozome.net/soybean). SimpleProbes were purchased from Roche Applied Sciences. The Lpa1 probe consists of 5’-Fluorescein-SPC-AGAAGAGAAAGCAAACGATCGA-GAAAGAA-C3 blocker; A > G mutation indicated by underline. The Lpa2 probe consists of Fluorescein-SPC-TTTGCTGCTAGTAAATCTCTCAATAGATAT-C3 blocker (reverse complement of CX1834; G > A mutation indicated by underline). Primers used to generate template for SimpleProbe genotyping assays were designed by aligning the Lpa1 and Lpa2 sequences and regions near the SNPs. Primers were selected to be as close as possible to the SNP while differing by at least three nucleotides between the two genes to specifically amplify either Lpa1 or Lpa2. Genotyping reactions used asymmetric PCR to generate additional single-stranded DNA to which the SimpleProbe could bind with less competition from the opposite amplification strand. Lpa1 genotyping reactions were performed with a 5:1 asymmetric mix of primers (5’-CCTACTCCTCCTCTGAAGAG-3’ at 0.074 μM final concentration, and 5’-TATGCTGCTGCCATGTGAAAG-3’ at 0.375 μM final concentration), whereas a 10:1 ratio was observed to work best for Lpa2 (5’-CTTAATCTCTGGAGATCAAGTTATGTG-3’ at 0.375 μM final concentration, and 5’-AAAAAGAGCACAATGTGAAGCTG-3’ at 0.0375 μM final concentration).

Reactions were performed in 20 μL containing template, primers, 0.2 μM final concentration of SimpleProbe, buffer (40 mM Tricine-KOH [pH 8.0], 16 mM KCl, 3.5 mM MgCl₂, 3.75 μg mL⁻¹ BSA, 200 μM dNTPs), 5% DMSO, and 0.2× Titanium Taq polymerase (BD Biosciences). Lpa1 genotyping reactions were performed using a Lightcycler 480 II real-time PCR instrument (Roche), using the following PCR parameters: 95°C for 5 min followed by 40 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 30 s, and then a melting curve from 55 to 70°C. The fluorescence was read after each cycle and every 0.1° during melting curve analysis. For Lpa2, the following PCR conditions were used: 95°C for 5 min followed by 45 cycles of 95°C for 30 s, 59°C for 30 s, 72°C for 30 s, and then a melting curve from 51 to 66°C. When DNA from CX1834 (lpa1-a, lpa2-a) or M766 (lpa2-b) is amplified with gene-specific primers and used in melting curve analysis with the SimpleProbe, a mismatch between the
SimpleProbe and the amplicon results in altered disassociation kinetics. Each genotype produced a characteristic melting profile, as measured by melting temperature (Tm) of the negative first derivative of the disappearance of fluorescent signal. Lpa1 (W82/5601T) gave a characteristic peak of 66.5°C, while the peak for lpa1-a (CX1834) was 63°C. The Lpa2 alleles from W82 and 5601T yielded a characteristic peak at 62.5°C, and lpa2-a (CX1834) gave a characteristic peak at 57°C and lpa2-b (M766) gave a characteristic peak at 58.5°C.

Certain samples were also examined with a DNA Engine Opticon 2 (MJ Research/Bio-Rad) using the same cycling conditions and peak identification methods described above. In general, there was <0.5°C of difference between results obtained on the two instruments.

Genotyping for Lpa1 and Lpa2 in the three populations was performed with the SimpleProbe assays as described or by direct sequencing of PCR products.

**Results and Discussion**

**Identification of Mutant Soybean MRP4 Homologs**

In a previous study, a mutation within a conserved region of a maize multidrug resistant protein class 4 gene (MRP4) was identified as causative for the low phytic acid phenotype (maize lpa1-1). In the same study, the authors created a gene-silencing construct containing a portion of a soybean homolog to the maize lpa1 gene. Expression of this silencing construct in transgenic soybean plants resulted in soybean seeds containing reduced phytic acid and an increase in seed inorganic phosphate (Shi et al., 2007).

Due to the known multiple polyploidization events for soybean (Shoemaker et al., 2006), we reasoned that while a single gene is responsible for the phenotype in maize, multiple genes would need to be silenced in soybean. To identify the genes silenced by this construct, we used the portion of the silencing construct homologous to the maize MRP4 gene in BLAST searches against the Glyma0 and Glyma1 whole-genome sequence (http://www.phytozome.net/soybean). BLAST searches identified two genes highly similar to the maize MRP4 gene, Glyma03 g32500 and Glyma19 g35230, on LG N/chr3 and LG L/chr19, respectively (Supplementary Fig. 1).

Expression of this silencing construct in transgenic soybean plants resulted in soybean seeds containing reduced phytic acid and an increase in seed inorganic phosphate (Shoemaker et al., 2006).

We used gene-specific PCR to amplify and sequence each of these two putative soybean Lpa genes from a normal phytic acid line, 5601T, and the low phytic acid donor CX1834 using the W82 (Bernard and Creemers, 1988) genomic sequence as reference sequence. For Glyma03 g32500 (Lpa1) we identified a nonsense mutation, A3828T relative to the start codon, in the CX1834 allele (Fig. 1), which results in truncated protein of 893 amino acid residues, compared with the reference protein containing 1492 amino acid residues (Fig. 2). For Glyma19 g35230 (Lpa2), we identified a missense mutation (Fig. 3), G4866A relative to the start codon, resulting in an R1039K substitution at a position conserved between Lpa1, Lpa2, maize MRP4, and Arabidopsis MRP5 proteins (Fig. 2). Because little is known about the amino acids required for function of MRP4/MRP5 proteins, we attempted to verify that this residue is indeed conserved. To that end, we used tBLASTn to search the entire expressed sequence tag (EST) collection for Viridiplantae MRP4/MRP5 matches, using a protein query in the region immediately proximate to the residue affected by the lpa2 mutation. In all ESTs identified by this tBLASTn search, the arginine residue is absolutely conserved (Fig. 4). We therefore conclude that this invariant position may be critical for protein function and/or stability. We also identified a number of intron and silent SNPs present in the 5601T Lpa1 allele, relative to the W82 reference sequence (data not shown). No differences were found between the Lpa2 W82 reference sequence and Lpa2 from 5601T. For clarity, we will henceforth refer to Glyma03 g32500 as Lpa1 and the nonsense mutation allele identified in CX1834 will be referred to as lpa1-a. Glyma19 g35230 will be referred to as Lpa2 and the mutant allele in CX1834 is designated lpa2-a.

**Development of Molecular Marker Assays**

Molecular marker assays were designed to distinguish lpa1-a and lpa2-a alleles from their wild-type counterparts. Because of the high degree of sequence similarity between the genes, we elected to use gene-specific PCR to generate templates to which a SimpleProbe is used to distinguish between the CX1834 and 5601T or wild-type alleles (Fig. 1 and 2). Both the lpa1-a and lpa2-a mutations are located in regions where the W82 reference
genomic sequence (http://www.phytozome.net/soybean) and 5601T sequences are identical. SimpleProbes were designed to be complementary to W82 allele (http://www.phytozome.net/soybean), and any mismatch between the SimpleProbe and target sequence is detected during melting peak analysis. SimpleProbes yield identification of both heterozygote and homozygote genotypes.

**Association Analysis for Inorganic Phosphate Phenotype and lpa1-a and lpa2-a Genotypes**

We examined two segregating populations derived originally from a cross between 5601T × CX1834. Population 1 was an F5:8 RIL population containing 144 individuals that had been previously characterized for inorganic phosphate content and Satt237 and Satt561 marker genotypes (Walker et al., 2006). We genotyped samples from this population for the lpa1-a and lpa2-a alleles, and the results indicated a complete association between the inheritance of homozygous lpa1-a and lpa2-a alleles and a high inorganic phosphate phenotype (Fig. 5). Both loci are required to manifest the increased inorganic phosphate phenotype, and even a single wild-type copy of either Lpa1 or Lpa2 resulted in a decrease in phosphorus levels from an average of 2008 μg P g−1 seed (lpa1-a/lpa1-a, lpa2-a/lpa2-a) to an average of 970 μg P g−1 (lpa1-a/lpa1-a, Lpa2/lpa2-a) or 804 μg P g−1 (lpa1-a/Lpa1, lpa2-a/lpa2-a). There were no obvious recombinants in this population between the SSR markers and the Lpa1 and Lpa2 genotype.

A second population derived from the 5601T × CX1834 cross was also available for association analysis. In this second population, a single F5:8 RIL derived

<table>
<thead>
<tr>
<th>lpa1</th>
<th>lpa1-a</th>
<th>lpa2</th>
<th>lpa2-a</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3644) TATCCTACTGCACTCTGTGAAGA</td>
<td>(3644) TATCCTACTGCACTCTGTGAAGA</td>
<td>(3201) TATCCTACTGCACTCTGTGAAGA</td>
<td>(3201) TATCCTACTGCACTCTGTGAAGA</td>
</tr>
</tbody>
</table>

Figure 1. Sequence of soybean low phytic acid (lpa) genes in lpa1-a SNP region and lpa1-a molecular marker assay allele discrimination. (A) Portion of the Lpa1 (Glyma03 g32250) sequence aligned with related sequences. Arrow indicates position of lpa1-a (CX1834) mutation (A3828T, relative to start codon). Black background with white text represents primers used for amplification for SimpleProbe analysis, and underline indicates region to which SimpleProbe corresponds. Gray background with black text indicates differences between Lpa1 and Lpa2. (B) Typical Lpa1 genotyping result with SimpleProbe on Lightcycler 480 II (Roche Applied Sciences). Lpa1 (W82/5601T) gives a characteristic peak of 66.5°C, and the peak for lpa1-a (CX1834) is 63°C. Heterozygotes show both peaks.
from an F₁ plant that was double heterozygous at Satt237 and Satt561 loci was used to produce a set of F₉:10 lines. Because we originally obtained a subset of this second population for independent experiments, we characterized only those lines that had been previously genotyped to be homozygous for the Satt237 and/or the Satt561 markers in the four possible homozygote SSR marker classes. We determined the inorganic phosphate levels and Lpa1 and Lpa2 genotype of individual seeds for a total of 120 F₉:10 NIL-derived samples. In this population, there were a small number of discrepancies between the Lpa1 genotype and the previously assigned Satt237 genotype, which we largely confirmed to be errors in the original Satt237 genotype. As in the first population, there was a complete association between an increased level of inorganic phosphate and homozygosity for the lpa1-a and lpa2-a mutations (Fig. 5). Furthermore, the second population also revealed that the high-phosphate phenotype is due to homozygosity for both lpa1-a and lpa2-a (average 2284 μg P g⁻¹ seed). Notably, a single copy of either Lpa1 or Lpa2 was observed to reduce inorganic phosphate concentration to levels not significantly different from double homozygote Lpa1/Lpa2 seeds (440 μg P g⁻¹ seed for lpa1-a/Lpa1 and lpa2-a/lpa2, 482 μg P g⁻¹ seed for lpa1-a/lpa1-a and lpa2-a/Lpa2, and 371 μg P g⁻¹ seed for Lpa1/Lpa1 and Lpa2/Lpa2).

Genetic Recombination between SSR Markers and Lpa Gene Loci

Breaking the linkage between the Lpa1 and Lpa2 genes and their closely associated SSR markers Satt237 and Satt561, respectively, would provide important insight into the relevance of the lpa1-a and lpa2-a alleles’ control of the low phytic acid trait. We screened all of our...
material from the second population for inorganic phosphate phenotype that had been originally genotyped as having the 5601T Satt237 alleles and the CX1834 Satt561 alleles. We identified two cases where recombination resulted in disconnect between the CX1834 lpa1-a allele and the SSR marker linked to low phytic acid in CX1834 (Satt237). We confirmed two individual plants where recombination resulted in an increased inorganic phosphate phenotype despite the presence of the 5601T allele of SSR marker Satt237. Although the two loci are physically extremely close to one another (Fig. 6), this...
result suggests a recombination occurred between the SSR marker and the $lpa1-a$ allele, breaking the linkage between the CX1834 allele of Satt237 and the $lpa1-a$ gene. We were unable to determine the exact breakage point where the recombination occurred. We were unable to identify similar recombinants for $lpa2-a$. Nevertheless, this result supports the hypothesis that the mutant allele of $lpa1-a$, in combination with $lpa2-a$, is responsible for the elevated inorganic phosphate phenotype.

**Independent Association of $lpa1-a$ and $lpa2-a$ with Phosphate Phenotype**

We also examined a portion of a third advanced population derived from a cross between CX1834 × S97-1688. Out of the 94 lines that were evaluated for inorganic phosphate phenotype, the six lines that possessed increased inorganic phosphate were genotyped for $Lpa1$ and $Lpa2$. As in the other two populations, without exception, all seeds displaying an increased inorganic phosphate phenotype were homozygous for both $lpa1-a$ and $lpa2-a$. The average inorganic phosphate content for the $lpa1-a$ and $lpa2-a$ mutant seeds was very similar to the other two populations we examined (2370 μg P g⁻¹ seed with a standard deviation of 592 μg P g⁻¹ seed).

**Candidate Gene Analysis of Low Phytic Acid Line M766**

During mutagenesis screening, Wilcox et al. (2000) obtained two independent low phytic acid soybean lines from mutagenesis of breeding line CX1515-4:M153, which was used in the creation of the CX1834 lines, and M766 (Wilcox et al., 2000). M153 possessed a slightly higher reduction in phytic acid compared with M766. For this reason, M766 has apparently not been utilized in breeding efforts or genetic analysis, to our knowledge. To address whether M766 possessed a novel mutation, we elected to sequence the $Lpa1$ and $Lpa2$ genes from low phytic acid soybean line M766. Upon sequencing $Lpa1$ from line M766, we noted only a single base change compared to the W82 reference sequence. A single SNP was identified, T5202A relative to start codon, located within intron 9. Because we do not currently have access to a segregating population containing the M766 allele, it is not clear if this SNP is causative of the high inorganic phosphate phenotype in M766. The M766 allele of $Lpa2$ was found to contain an independent nonsense mutation (Fig. 3; A4864T, relative to the start codon). This nonsense mutation in $Lpa2$ results in premature termination at residue 1038 (compared with 1316 in the full-length protein). We have designated the M766 allele $lpa2-b$. Fortuitously, the two $Lpa2$ mutant alleles, $lpa2-a$ and $lpa2-b$, are readily distinguishable from each other and wild-type $Lpa2$ with the same SimpleProbe assay (Fig. 2). Neither the $lpa1-a$ nor $lpa2-a$ mutations are present in M766.

**Origin of Soybean $lpa$ Mutations**

It has been previously suggested that one, or potentially both, soybean $lpa$ mutations may have existed before
the ethyl methanesulfonate (EMS) mutagenesis treatment (Walker et al., 2006). The original parental line (CX1515-4) used to develop M153 and M766 is no longer available (R. Nelson, personal communication, 2009). As a result, we were unable to directly address this question. Nevertheless, we investigated a subset of the North American ancestral soybean lines, which are thought to have contributed a significant majority (>80%) (Gizlice et al., 1996) to the genome of all North American cultivars (PI 506582 Asahi 60, PI 592941 Liao dou No. 10, PI 341246 CNS, PI 506924 Kosaka zairai (Kurobeso), PI 54610, PI 81041 Kuro Daizu, PI 548485 Roanoke, PI 548406 Richland). The lpa1-a, lpa2-a, or lpa2-b alleles were not present in any of these soybean lines.

In addition, M766 (an M3 line derived from EMS treatment of CX1515-4 seed) and M153 were produced by the same CX1515-4 mutagenesis experiment, and the M766 and CX1834 alleles do not possess any SNPs in common; each allele is unique. We therefore conclude that the lpa1-a, lpa2-a, and lpa2-b mutations likely arose as a direct result of the mutagenesis treatment. Although it has been thought to be extremely unlikely to randomly mutate two independent genes required to express a single phenotype, the screening power of the simple
inorganic phosphate assay was very efficient (Wilcox et al., 2000). In addition, we have confirmed a similar two independent gene mutagenesis when screening for low linolenic acid levels in the oil fraction of soybean seeds where neither mutation was found in ancestral lines (Bilyeu et al., 2005; J. Gillman, unpublished, 2009).

**Conclusions**

Fine mapping and analysis of multiple lpa mutants in maize (Shi et al., 2007) resulted in the identification of mutations within an MRP4/MRP5 gene as directly responsible for seed phytic acid levels. The specific mode of action for the reduction of seed phytic acid remains somewhat ambiguous. Although the members of the Arabidopsis MRP5 gene family possess ABC domains and are thought to act as vacuolar transporters involved in guard cell function (reviewed in Klein et al. 2006), it is not immediately obvious how two nonfunctional MRP4/MRP5 genes would lead to a decrease in seed phytic acid levels and a concomitant increase in inorganic phosphate. One intriguing hypothesis is that Lpa1 and Lpa2 may be involved in the movement of free phosphorus, phytate, or an inositol phosphate intermediate, between subcellular compartments within the developing soybean seed.

Soybean differs from maize in that two recessive MRP4/MRP5 mutant loci are required for the low phytic acid phenotype, rather than the single recessive mutant locus identified in maize. The identification of a nonsense mutation within a candidate lpa1 homolog, Glyma03 g32500 (Maroof et al., 2009), suggested that the nonsense mutation on LG N/chr3 could play a role in the molecular basis for the low phytic acid phenotype in soybean. We independently discovered the nonsense mutation within the CX1834 allele of Glyma03 g32500. We have also identified a novel missense mutation in a conserved portion of the other lpa1 homolog, Glyma19 g35230 in CX1834. Together, the results provide clear evidence that two soybean lpa homologs control the low phytic acid phenotype.

We developed high-throughput molecular marker assays to directly select for the alleles that control the soybean low phytic acid phenotype. We have utilized these molecular marker assays to reveal a perfect association between the recessive mutations in both soybean genes homologous to the maize lpa1 gene and the low phytic acid phenotype. Lastly, we report the identification of the novel lpa2-b allele in M766, a line that, to our knowledge, has not been utilized in any soybean breeding project. Our results suggest the next step for soybean low phytic acid breeding efforts may lie in the combination of the nonsense lpa2-b allele from M766 with the nonsense lpa1-a allele from CX1834. Such a combination may produce soybeans with even lower levels of phytic acid and increased available phosphate levels, yielding enhanced nutritional value for food and feed.

**Acknowledgments**

The authors appreciate the excellent technical assistance of Ashley Tetlow. This research was supported by the USDA-Agricultural Research Service and the United Soybean Board.

**References**


Bilyeu, K.D., P. Zeng, P. Coello, Z.J. Zhang, H.B. Krishnan, A. Bailey, P.R. Beuselinck, and J.C. Polacco. 2008. Quantitative conversion of...


