The Genetic Basis of Phytate, Oligosaccharide Content, and Emergence in Soybean

Natasha Marie Glover

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Crop and Soil Environmental Sciences

M.A. Saghai Maroof, Committee Chair
Elizabeth A. Grabau
Richard E. Veilleux
John G. Jelesko
Ina Hoeschele

June 27, 2011
Blacksburg, VA

Keywords: sucrose, stachyose, raffinose, MIPS, MRP

The Genetic Basis of Phytate, Oligosaccharide Content, and Emergence in Soybean

Natasha Marie Glover

ABSTRACT

Soybean [*Glycine max* (L.) Merr] is one of the U.S.ôs most economically important crops due to the protein and oil content of seeds. The major storage form of phosphorus in soybean seeds is found in the form of phytate, but because of its negative nutritional and environmental impacts, seed phytate and raffinosaccharide content have been a recent focus of breeders and molecular geneticists. The soybean line CX1834 is a low phytate mutant known to have two low phytate QTLs on linkage groups (LGs) L and N. The first objective of this research was to determine the genetic basis of the low phytate trait in CX1834. By using the whole genome sequence, we identified two candidate multidrug resistance-associated (MRP) ABC transporter genes. Sequencing the genes from CX1834 and comparing them to the reference genome sequence revealed a single nucleotide polymorphism (SNP) in the MRP gene located on LG N (causing a stop codon), and a SNP mutation in the MRP gene located on LG L (causing an amino acid change from arginine to lysine).

One major concern with low phytate soybeans is the low seedling emergence. The second objective was to undertake a population-wide study of emergence in the recombinant inbred population CX1834 x V99-3337, over two years and two locations. We found a positive correlation between phytate level and emergence, and that variation among year, location, genotypic class, year x genotypic class, and year x location interactions were significantly affecting emergence.

V99-5089, in addition to being low phytate, has high sucrose and low raffinosaccharide content. This phenotype of V99-5089 has been previously determined to be due to a SNP

mutation in its *myo*-inositol phosphate synthase (MIPS) gene located on LG B1. The third objective was to use the recombinant inbred population derived from CX1834 x V99-5089 to observe the combinations of all three mutations to see how the different alleles impact phytate and raffinosaccharide content. The individuals with all three mutations, as well as those with the two MRP mutations together had lower phytate than the other genotypic classes. However, these lines (all three mutations) had unexpectedly high stachyose.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude to my advisor, Dr. M.A. Saghai Maroof for his guidance and support throughout my undergraduate and graduate studies. I would like to thank my committee members, Dr. Elizabeth Grabau, Dr. Ina Hoeschele, Dr. John Jelesko, and Dr. Richard Veilleux, for their advice and guidance over the years. I would like to especially thank Dr. Ruslan Biyashev for everything he has taught me the past nine years.

I also would like to show my appreciation for the Multicultural Academic Opportunities Program (MAOP) at Virginia Tech for encouraging me to pursue an advanced degree, for mentoring me, and for providing financial support since freshman year. I would like to thank those who have helped advise this research, including Dr. Chao Shang and Dr. Glenn Buss. I am sincerely appreciative of the help and friendship provided by my graduate and postdoc colleagues: Dr. Tatiana Boluarte, Brian Bowman, Dr. Star Gao, Dr. Juan Ruiz-Rojas, Christin Kastl, Dr. Jafar Mammadov, Dr. Jeff Skoneczka, and Dr. Dominic Tucker. I would also like to thank the faculty and staff in the department of Crop and Soil Environmental Sciences for the support they have offered me from the beginning of my college education.

There are countless others who have assisted and encouraged me over the years, but I would above all like to thank my family for their love and support. I am extremely grateful for my parents, Debra and Ronald Glover.

LIST OF FIGURES

LIST OF TABLES

I. LOW PHYTATE AND ITS ROLE IN AGRICULTURE (LITERATURE REVIEW)

Abbreviations: ABC, ATP-binding cassette; AFLP, amplified fragment length polymorphism; BAC, bacterial artificial chromosome; EMS, ethyl methanesulfonate; EST, expressed sequence tag; Ins, inositol; LP, low phytate; *lpa*, low phytic acid; MIPS, *myo*-inositol phosphate synthase; MRP, multidrug resistance-associated protein; ORF, open reading frame; P, phosphorus; PA, phytic acid; Pi, inorganic phosphorus; PSV, protein storage vacuole; RIL, recombinant inbred line; SNP, single nucleotide polymorphism; SSR, simple sequence repeat

PHYTIC ACID

Phytic acid (PA or $InsP_6$), otherwise known as *myo*-inositol (1,2,3,4,5,6)-hexa*kis*phosphate, is the major form of phosphorus (P) in seeds, usually representing about 75% of total seed phosphorus (Raboy, 2001). The chemical structure of phytic acid (Figure 1.1) consists of a six-membered ring with six phosphate groups attached to each carbon in the ring. A large portion of phosphorus taken up from the soil by crops eventually gets translocated to the seeds, where it is converted to phytic acid. Phytic acid is widespread in eukaryotes and has been shown to have a role in numerous cellular functions, including signal transduction, DNA repair, RNA transfer from the nucleus, ATP regeneration, control of guard cells, and phosphorus and mineral storage (Raboy, 2009). In plants, phytic acid may play a role in the regulation of K^+ and Ca^{2+} flux of guard cells. This is important in controlling turgor pressure and transpiration of plant cells (Lemtiri-Chlieh et al., 2003). Most phytic acid is deposited in storage vacuoles (globoids) as phytin salts, which include complexes with mineral cations including zinc, iron, potassium, magnesium, and calcium (Lott et al., 1995). In its salt form, phytic acid is known as phytate, and from here on will be referred to as phytate. Phytate is broken down by phytase enzymes found within the plant during germination, and subsequently P, *myo*-inositol, and minerals are released to be used by the seedling.

PHYTATE BIOCHEMICAL PATHWAY

The phytate pathway is not completely understood, but low phytic acid (*lpa*) mutant plants can be used to help understand which enzymes are involved in phytic acid biosynthesis. Phytic acid can be synthesized by one of two pathways in plants: a lipid-dependent or lipid-independent pathway. The lipid-dependent pathway comes from the hydrolysis of PI(4,5)P2 by phospholipase C. The lipid-independent pathway consists of the sequential phosphorylation of InsP3 or inositol (Stevenson-Paulik et al., 2005). Phytic acid biosynthesis generally proceeds through the lipid-independent pathways in the seeds of cereals and legumes (Rasmussen et al., 2010).

The first step of the phytic acid pathway is the conversion of glucose-6-phosphate to InsP3 by *myo*-inositol phosphate synthase (MIPS). This reaction is the first committed step in *myo*-inositol production, and is the de novo source of inositol in plants (Loewus and Murthy, 2000). MIPS catalyzes the conversion of glucose-6-phosphate to inositol 1-phosphate by an oxidation, enolization, condensation, and reduction reaction. Then, inositol-1-phosphate is dephosphorylated to inositol by the enzyme IMPase. Phytic acid is subsequently produced by the stepwise phosphorylation of *myo*-inositol by a number of different kinases (Figure 1.2).

Because phytic acid chelates mineral cations including calcium, iron, and zinc, these mixed salts are often excreted by non-ruminant animals such as humans, swine, poultry, and fish (Brinch-Pedersen et al., 2002). In humans, this could cause iron and zinc deficiencies in developing areas of the world where the country relies on a sole staple crop for food (Raboy, 2001). Between 29-65% of the total soil P is organic P (Dalai, 1977; Harrison, 1987). A differential stabilization exists for organic P compounds in soils, where the compound is more strongly adsorbed to the soil as the amount of phosphate groups increases (Celi et al., 1999).

Phytic acid has six phosphate groups, and therefore a high charge density, causing it to bind strongly to the soil (Anderson and Arlidge, 1962; Celi et al., 1999). Thus, phytic acid is stabilized and accumulates in soils due to this strong interaction, and the majority (up to 90%) of the organic P is phytic acid in most soils (Harrison, 1987).

The fact that phytic acid stabilizes and accumulates in most soils may indicate that is biounavailable for uptake by plants, possibly due to its inaccessibility in the soil matrix. However, phytic acid can be used for uptake in several species in environments where P is limited (Corona et al., 1996; Kroehler and Linkins, 1991; Li et al., 1997). The large amount of inositol phosphates present in soil suggest there is a potential of transfer due to erosion, although there is little direct evidence showing that inositol phosphates in the soil transfer to runoff water (Turner et al., 2002). The majority of the P in animal feed is phytic acid, but because monogastric animals lack endogenous phytase, the enzyme that breaks down phytic acid, this P is unavailable (Morse et al., 1992). To meet the animalôs dietary requirements inorganic P (Pi) supplements can be used, but the uptake is inefficient, with 70% of the total P excreted. The manure is often used as agricultural fertilizer and applied to land, increasing the potential for P transfer to runoff water (Haygarth et al., 1998). To increase P uptake efficiency and the bioavailability of certain minerals, animal feed can be supplemented with phytase, for the breakdown of phytic acid in the animalôs gut (Lei et al., 1993; Poulsen, 2000). Although phytase supplements can enhance an animalôs diet, it may have undesirable environmental effects. Inositol phosphates found in manure are unlikely to be transferred to runoff because they are strongly bound to the soil, but by converting phytic acid to inorganic P, phytase supplements may actually increase the risk of P transfer to runoff waters (Turner et al., 2002).

Various feeding trials have shown that low phytate grain can increase the available P to animals, satisfying more of their dietary requirement for P (Ertl et al., 1998). In turn, the animals excrete less P, resulting in less P runoff into bodies of water. Human studies have shown that iron, zinc, and calcium are 35-50% more available in low phytate food (Raboy, 2007). As mentioned earlier, because of the high proportion of excreted P in monogastric animals such as poultry, swine, and fish, the animalôs nutritional requirement for P is usually supplemented with Pi or phytase supplements. These phytase supplements break down phytic acid in the feed, and release up to 50% of the phosphorus, making it available to the animal. This reduces the environmental consequences, however it may also contribute significantly to feed costs (Cromwell et al., 1995; Raboy, 2001).

LOW PHYTATE IN PLANTS

In crops, seeds normally contain between 3.0 and 8.0 mg of P per gram of dry weight, and of this total P, 65-75% is phytate. About 5% of the remaining total P is Pi, and less than 10% are lower-order *myo*-inositol phosphates such as InsP5, InsP4, etc. (Raboy, 2007). Due to the indigestibility of phytate in humans and animals and its environmental consequences, low phytate varieties of crops are beneficial (Raboy, 2007). However, few, if any, naturally occurring varieties of low phytate crops are known to exist. Low phytate plants are usually developed by mutagenesis, and seeds homozygous for a particular *lpa* allele usually have a reduction in seed phytic acid of 30-90%. This was first demonstrated in the 1990s with the production of two low phytate maize mutants, *lpa1-1* and *lpa2-1*(Raboy and Gerbasi, 1996). The reduction of phytic acid is generally coupled with an increase in seed Pi, to maintain a constant total P content of the seed. Due to this inverse relationship between phytate and Pi, a Pi assay can be used to screen for low phytate individuals. Within a cell, phytate is found in globoids, which are discrete globular inclusion bodies found in protein storage vacuoles (PSVs). In cereals, these phytate-containing PSVs can be found in the aleurone and germ tissues, whereas in legumes they are located within the germ and cotyledonary tissues (O'Dell et al., 1972).

As mentioned earlier, the phytic acid biosynthesis pathway involves a number of substrates and enzymes (Figure 1.2). Numerous low phytate mutants have mutations in one or more of the enzymes involved in the pathway, presumably decreasing the accumulation of *myo*-inositol and resulting in low seed phytate. When used in feeding trials, these low phytate mutants can lower the amount of P in animal waste and improve iron, zinc, and calcium nutrition in humans and animals (Raboy, 2007). The following section discusses the current status of low phytate plants, including the forward/reverse genetics approaches to developing low phytate

crops, the characterization of the low phytate crops, and the description of any causal genes if they have been identified. Maize, wheat, barley, rice, and soybean will be discussed, as more research has been conducted on these crops due to their economic importance. However, there have been numerous studies in plants such as *Arabidopsis thaliana*, common bean, *Brassica rapa*, etc. that will not be discussed here (Campion et al., 2009; Fileppi et al., 2010; Kim and Tai, 2010; Nagy et al., 2009; Panzeri et al., 2011; Zhao et al., 2008).

Low phytate in maize

Development of Low Phytate Mutants

Maize (*Zea mays*) has four known low phytate mutants: *lpa1*, *lpa2*, *lpa3*, and *lpa241* (Pilu et al., 2003; Raboy and Gerbasi, 1996). Different alleles of a particular *lpa* gene are denoted with a suffix such as ñ-1,ò ñ-2,ò etc. *Lpa1-1* and *lpa2-1* are both mutants that were developed by treating pollen with ethyl methanesulfonate (EMS), and then the pollen was used to pollinate silks of different plants, resulting in heterozygous M_1 seeds. The M_1 plants were selfed to produce M_2 seeds, which were screened for the low phytate phenotype (Raboy and Gerbasi, 1996; Raboy et al., 2000). Seeds homozygous for *lpa1-1* and *lpa2-1* had a reduction in phytate by ~66% and ~50%, respectively. The low phytic acid phenotype observed in *lpa1-1* and *lpa2-1* was accompanied by an increase in Pi, and this counterbalance therefore caused no effect on total seed P (Raboy et al., 2000). The authors hypothesized that *lpa1-1* was due to a mutation in the beginning of the phytic acid pathway, because of an observed reduction of all inositol phosphate species in seeds of *lpa1-1* mutants (Raboy et al., 2000). The authors also hypothesized that *lpa2-1* was due to a mutation in the later part of the phytic acid pathway, because although total inositol phosphates were reduced, there was an increase in other inositol phosphates not normally represented in seeds (InsP3, InsP4, InsP5).

Lpa1-1 and *lpa2-1* were genetically mapped to two loci on chromosome 1S in maize. *Lpa1-1* was mapped near a MIPS gene on chromosome 1S. Shukla et al. (2004) determined that MIPS enzyme activity and gene expression was decreased by 2-3 fold in mutant kernels. However, upon sequencing the MIPS gene, there were no sequence differences detected between the wild-type and mutant. There was a reduced yield of about 6% in *lpa1-1* near-isogenic lines, and a reduced seed dry weight for all lines.

The maize *lpa3* mutant was developed by screening F_2 seeds derived from *Mutator* (*Mu*) stocks (Shi et al., 2005). *Mu* stocks are maize lines that contain multiple *Mu* transposons, thereby causing a high mutation rate (Planckaert and Walbot, 1989). One high Pi/low phytate line was identified, and inheritance experiments showed that the low phytate allele was due to a single-locus recessive mutation, and that *lpa3* was not allelic to any other known mutants. After the development of these maize *lpa* mutants, several studies were conducted in order to identify the genes responsible for the *lpa* phenotype.

Characterization of Lpa2

Shi et al. (2003) used a reverse genetics approach to identify the gene responsible for the *lpa2* low phytate phenotype in maize. As previously stated, there are many kinase enzymes involved in the *myo*-inositol biosynthesis pathway, and disrupting one or a number of them could have implications for phytate levels in seeds. Shi et al. (2003) investigated inositol phosphate kinase genes in maize as a possible cause for low phytate in the *lpa2* mutant. Putative maize inositol phosphate kinase genes were identified by searching an EST database of maize cDNAs with sequence similarity to *Arabidopsis* and human Ins(1,3,4)P3 5/6-kinase genes. Primers were designed to amplify and clone the identified corresponding maize genomic sequence, designated ó*ZmIpk*ô (*Zea mays* inositol phosphate kinase). *ZmIpk* was shown to have multiple inositol

phosphate kinase activites and phosphorylate several inositol monophosphates and polyphosphates. In order to determine the function of *ZmIpk,* the authors used a *Mu* insertion to disrupt the *ZmIpk* gene. The result of mutagenesis was a recessive loss of function allele, causing a 30% reduction in phytic acid of kernels. Due to the similar phenotype of the *ZmIpk Mu* insertion knockout and the maize low phytate line *lpa2*, the authors conducted an allelism test to determine if *lpa2* was the same gene as *ZmIpk*. Homozygous *ZmIpk* mutants were crossed with homozygous *lpa2* mutants, and all the F_1 seeds had a low phytate phenotype. This indicates that the low phytate displayed in *lpa2* is most likely due to a mutation in the *ZmIpk* gene. After amplification and sequencing, it was determined that *lpa2-2* allele was due to a point mutation of a C to T, introducing a stop codon at amino acid position 35. *Lpa2-1* was determined to be a rearrangement of the genomic sequence (Shi et al., 2003).

Characterization of Lpa3

A novel *lpa* mutation (*lpa3*) causing a low phytic acid phenotype was reported by Shi et al. (2005). This maize *lpa3* mutant was isolated by screening ~40,000 F_2 families derived from *Mu* stocks for the high Pi phenotype. From the screening, two *lpa3* lines were isolated, confirmed to be allelic, and denoted as *lpa3-1* and *lpa3-2*. Both *lpa3* mutant lines had increased levels of Pi and *myo*-inositol, but no increase in mono and poly-phosphate inositol intermediates, thereby differentiating them from *lpa2* mutants (Shi et al., 2005). In order to determine which gene had been mutated, the authors used a PCR-based method to determine the sequence surrounding the *Mu* insertion site. The *Mu* insertion site sequence was used to search the maize genome sequence for the disrupted gene. In a segregating population, individuals with a *Mu*-insertion corresponded with the low phytate phenotype. In addition, a reverse-genetics approach was taken to identify other lines with *Mu* mutations in the same *lpa3* gene. From these

experiments, two additional lines were identified with the same low phytate phenotype and with mutations in the *lpa3* gene. The *lpa3* gene was cloned, sequenced, and found to have sequence homology to a *myo*-inositol kinase gene (Shi et al., 2005).

Characterization of Lpa1

Homozygous low phytate F_2 lines derived from *Mu* stocks were crossed with a homozygous *lpa1-1* line to find novel insertion mutations in the *lpa1* gene identified by Raboy and Gerbasi (Raboy and Gerbasi, 1996; Shi et al., 2003). Three independent *lpa* mutants were identified with *Mu* inserted into the gene. In one of the mutants, phytic acid was reduced by 93%. The authors were initially unable to clone the *lpa1* gene, but instead identified a *Mu* insertion in a transcriptional activator protein gene located ~1 cM from *lpa1*. This transcriptional activator gene was used as a marker to identify a BAC clone. A map-based cloning strategy was used to develop additional markers from BAC-end sequences for fine mapping. Genetic analysis revealed that *lpa1* was between two markers, and *lpa1* was located on one or both of two overlapping clones. Sequencing of these two clones (~300 kb total) revealed nine ORFs. After searching EST databases for putative proteins encoded by the ORFs, the authors then searched for *Mu* insertions in the sequences from the mutant lines. They found that the *lpa1* gene encodes a multidrug resistance-associated protein (MRP) ATP-binding cassette (ABC) transporter due to the identification of *Mu* insertions in the *lpa1* mutant lines. In addition, in a segregating population of 792 individuals, the *Mu* mutation cosegregated perfectly with the low phytate phenotype. In the maize *lpa1-1* mutant line, there was a point (missense) mutation resulting in an alanine to valine change at the 1432 amino acid, which is normally a conserved amino acid (Shi et al., 2007).

RT-PCR, Northern analysis, and signature sequencing analysis were performed on *lpa1* transcripts in order to characterize the expression pattern of the *lpa1* gene. The *lpa1* gene is expressed in seeds, endosperm, and embryos at various days after pollination, and analysis of ESTs showed expression in roots, leaves, stalks, tassels, silks, developing seeds, and seedlings.

Upon cloning the *lpa1* gene, the authors used gene-silencing constructs to suppress *lpa1* gene expression (Shi et al., 2007). In maize, the embryos are the site of phytic acid biosynthesis and accumulation, therefore the embryo-specific *Ole* and *Glb* promoters were chosen to make gene-silencing constructs. A super binary maize transformation vector was used, and the construct contained a 5ô-end fragment of the MRP4 cDNA. Two types of gene-silencing constructs were made: one with the *Ole* promoter, and one with the *Glb* promoter. Transformants were produced by *Agrobacterium*-mediated transformation, and 80% of the transformants had a high Pi phenotype. Those transformants with the *Ole* promoter construct produced T_1 seeds with a 68-87% decrease in seed phytic acid, and those with the *Glb* promoter construct had T_1 seeds with a decrease in phytic acid by 32-75%. There was no significant difference between transformed and non-transformed seed weights. In order to see if MRP genes could be a means of reducing phytate in other crops, the soybean homolog of the maize MRP was silenced using a gene-silencing construct controlled by the soybean promoter Kunitz trypsin inhibitor 3 (KTI3). Transformation was done by particle bombardment of somatic embryos, and 14 out of the 23 independent transformants had increased Pi levels. There was a 37-90% reduction of phytic acid in transformed T_1 soybean seeds (Shi et al., 2007), indicating that silenced soybean MRP genes have the potential to reduce phytate.

Characterization of Lpa241

Lpa241 is another maize low phytate mutant, generated by EMS-treatment of pollen (Pilu et al., 2003). Six hundred M_2 families were screened for the high Pi phenotype, and one non-lethal mutant was isolated. This mutant, *lpa241*, had a 90% decrease in phytate, a 10-fold increase in Pi, and the total P content was unchanged. Although there was no significant change in dry weight of the *lpa241* seeds, there was approximately a 30% decrease in seed germination. Genetic tests based on the high Pi phenotype were done to determine the inheritance of the *lpa241* allele: A 3:1 segregation ratio of low Pi to high Pi was observed in the F_2 generation and a 1:1 segregation ratio of low Pi to high Pi in a test cross. These results indicated that *lpa241* is a single, recessive gene. Biochemical characterization determined that there was no accumulation of other inositol phosphates, suggesting a disruption in the first step of the pathway-- the reaction catalyzed by MIPS. Upon comparing the expression of the MIPS gene between the *lpa241* mutant and a wild-type, the authors determined that MIPS expression was ñremarkably weakerò in the shoot and kernel tissues of the low phytate mutant. Although no direct evidence was presented that MIPS is the cause of low phytate, the data suggestedô that MIPS is a good candidate gene and should be explored further (Pilu et al., 2003).

Low Phytate Wheat

Wheat (*Triticum aestivum*) is the staple food for 35% of the world population. Due to its chelation of mineral cations, a decrease in phytate would improve the bioavailability of these minerals and thus the nutritional value of the crop. In wheat and barley, phytate globoids are located mainly in the aleurone layer (O'Dell et al., 1972). In wheat, PSVs contain only matrix proteins and phytate globoids (Morrison et al., 1975), but are lacking the lattice structured

crystalloids (Lott, 1980). This makes the cells easily susceptible to leaching phytate into the environment from only a ñminor rupture of the aleurone cellsò (Antoine et al., 2004).

Several non-genetic strategies have been attempted to reduce the phytate level in cereals. For example, milling has been considered, but this removes not only the phytate, but also other key minerals. Soaking seeds in an aqueous solution can reduce phytate content, but comes with a loss of minerals, proteins and vitamins (Hurrell, 2004). The efficiency of the different methods for reducing phytate varies among wheat varieties. Due to these downfalls, reducing phytate genetically may be the most appropriate means. A transgenic approach has been used to introduce the *Aspergillus niger* phytase gene into wheat immature embryos and was shown to increase phytate degradation (Holm et al., 2002), although food and feed studies have not been reported yet. Several studies have evaluated large, diverse collections of wheat cultivars and mutants that identified a significant range in phytate content evaluated over multiple environments. This indicates that it is possible to develop low phytate wheat cultivars by traditional breeding (Ficco et al., 2009; Khan, 2007).

In order to develop low phytate mutants, wheat seeds were mutagenized with EMS by Gutteri et al. (2004). M_3 seeds were phenotyped for high Pi, and one nonlethal mutant was identified, designated as Js-12-*LPA*. Phytate was reduced by about 25% compared to wild-type wheat. In the low phytate mutant lines, P was increased in the endosperm and decreased in the bran. An inheritance study suggests low phytate is controlled by two or more genes in Js-12-*LPA*. Js-12-*LPA* was described as ñagronomically unacceptable,ò due to its lowered yield, and weak, reduced stature (Guttieri et al., 2004). Later research by the same authors evaluated the agronomic characteristics of the same low phytate wheat mutant. This was done by backcrossing the low phytate genotypes into several genetic backgrounds. Deleterious effects of being low

phytate were not consistent over three genetic backgrounds. This suggests that it may be possible to breed for low phytate wheat with improved agronomic traits (Guttieri et al., 2006).

Low Phytate in Barley

About 60-70% of barleyôs (*Hordeum vulgare*) seed total P exists as phytate (Larson et al., 1998). There are over twenty barley *lpa* mutants with at least six distinct mutant loci (Oliver, 2009). Four barley *lpa* mutations have currently been documented: *lpa1-1*, *lpa2-1*, *lpa3-1*, and M955. The inheritance and map positions of *lpa1-1* and *lpa2-1* are described by Larson et al. (1998). These two barley mutations were created by treating seeds of the variety ôHarringtonô with the mutagenesis agent sodium azide and screening for the low phytate phenotype. The *lpa1-1* mutation causes a decrease in phytate P and a subsequent molar-equivalent increase in Pi. The *lpa2-1* mutation results in a decrease in phytate but no molar equivalent increase in Pi. However, there is a 3-4 fold increase in Pi for *lpa2-1* compared to normal kernels. With *lpa1-1* there is a 5-10 fold increase of Pi (Larson et al., 1998). Mutants were selfed to create homozygous lines for *lpa1-1* and *lpa2-1*. These homozygous mutants were used as parents, and crossed with non-mutants to create F_2 mapping populations segregating for the mutation. *Lpa1-1* was mapped to chromosome 2H, and *lpa2-1* was mapped to chromosome 7H. In feeding studies, chicks fed low phytate barley, compared to chicks on a normal phytate diet, excreted 33% less P and displayed better growth and bone structure (Jang et al., 2003). Other studies have been conducted in pigs and rats with the similar results (Veum et al., 2002).

Characterization of Lpa1

In order to investigate the actual gene that corresponds to barley *lpa1* mutation, MIPS genes were considered as candidate genes, and were mapped. Their map positions were

compared to the map positions of barley *lpa1* (Larson and Raboy, 1999). As stated above, barley *lpa1-1* was mapped to chromosome 2H (Larson et al., 1998). In maize, seven sequences similar to MIPS were mapped to chromosomes 1S, 4L, 5S, 6S, 8L, 9S, and 9L. The position of the MIPS sequence on chromosome 1S correlates to the position of maize *lpa1-1*, supporting the idea that *lpa1-1* is due to a mutation in MIPS. However, in barley, only one copy of MIPS was detected, which mapped to chromosome 4H. This does not coincide with *lpa1* located on chromosome 2H, even though there is synteny between maize chromosome 1S and barley chromosomes 4H and 2H (Larson and Raboy, 1999).

Characterization of M955

Barley mutant M955 has over a 90% reduction in seed phytate compared to the wild-type (Bregitzer and Raboy, 2006). In order to investigate candidate genes for the genetic basis of M955, Bowen et al. (2007) used oligonucleotide microarrays to assay for differential gene expression between M955 and its corresponding nonmutant with the wild type allele. The results indicated a ñconsistent and substantialò difference in expression for 38 genes. These genes are involved in various cellular processes, however no differential expression was observed for genes known to be directly involved in phytate biosynthesis. Nevertheless, genes involved in carbohydrate and cell wall metabolism, cytokinin and ethylene signaling, and transport functions were differentially expressed, indicating that there is a relationship between phytate and oligosaccharide metabolism (Bowen et al., 2007).

Low Phytic Acid in Rice

Rice (*Oryza sativa*) is the major food crop for over 50% of Asians, and is a staple for almost two billion people (Liu et al., 2007). In 2000, Larson et al. isolated a non-lethal low

phytate rice mutant and genetically mapped its *lpa1* loci, along with a rice MIPS gene. Low phytate rice mutants were isolated by gamma-irradiating M_1 seeds, planting them, and screening the M_2-derived M_3 seeds for the high Pi/low phytate phenotype. Two putative low phytate mutants were identified, named óOrion M2-70ô and óKaybonnet M2-2045.ô However, the Orion M2-70 mutant was shown to be lethal, producing non-viable seeds in the M_3 and M_4 generation. Kaybonnet M2 was shown to have a non-lethal and heritable mutation named low phytate 1-1. *Lpa1-1* was determined to be a recessive allele at a single gene. Genetic mapping was conducted using the intersubspecific mapping population Kaybonnet x óZHE733.ô *Lpa1-1* was mapped on rice chromosome 2 using AFLP analysis. In addition, the rice MIPS gene was mapped to chromosome 3, where it is orthologous to previously mapped MIPS on chromosome 1S in maize, and chromosome 4 in Triticeae. Due to the location of the rice *lpa1* gene differing to that of MIPS, MIPS was ruled out as a candidate gene.

Characterization of Lpa1

In order to further examine the *lpa1* locus, previously mapped to chromosome 2 in rice, Andaya and Tai (2005) fine mapped *lpa1* in an RIL population. Using SSRs, they mapped the locus between two markers ~135 kb apart from each other. By developing markers from DNA sequence, they were further able to fine map the locus to a 47 kb region. This region contained eight ORFs and opened the door for cloning and further characterization of the gene encoding *lpa1* (Andaya and Tai, 2005).

In hopes of developing and characterizing novel *lpa* rice mutant lines to be used in breeding, Liu et al. (2007) generated eight *lpa* mutant lines in both *indica* and *japonica* subspecies. These mutants were created by either gamma rays or NaN_3 treatment. The non-lethal mutants had reduced phytate levels by 34-64%. From crossing studies, it was shown that

all the *lpa* mutations were single, recessive alleles. The authors mapped two of the mutations to chromosomes 2 and 3. One of the mutations mapped to chromosome 2 between two markers where the rice *lpa1* locus was previously mapped by (Andaya and Tai, 2005). The other *lpa* mutation from this study was mapped to chromosome 3, and is most likely a homolog of the maize *lpa3* (a *myo*-inositol kinase) gene, although more cloning/biochemical studies are needed to prove this. In 2009, Xu et al. mapped the low phytate allele from Os-*Lpa*-XS-110-2, a homozygous non-lethal mutant, to chromosome 3, a region containing the rice ortholog (OsMRP5) of the maize *lpa1* gene. Sequencing determined that there was a single base mutation of a C/G to T/A in the sixth exon, which reduces seed phytate by ~20% (2009).

Kim et al. (2008) carried out a similar study in which they isolated two independent, single gene, recessive *lpa* mutants: N15-186, and N15-375. Phytate was reduced by 75% and 43% in N15-186 and N15-375, respectively. Upon further characterization of the N15-186 mutant, it was revealed that its phytate, inositol monophospate, Pi, and *myo*-inositol levels were similar to those described for the maize *lpa3* mutant. Mapping revealed the mutation to be located at the same region as found by Liu et al. (2007) on chromosomes 2 and 3. Sequencing revealed a single base mutation in exon 1 of the *myo*-inositol kinase (MIK) gene, the rice homolog of maize *lpa3*. This mutation is a nonsense mutation of a C/G to T/A (Kim et al., 2008).

Low Phytate in Soybean

Soybeans (*Glycine max* (L.) Merr) are an important crop in the U.S. and worldwide. They are globally the number one and number two largest sources of protein feed and vegetable oil, respectively. The U.S. is the worldôs leading soybean producer and exporter, with U.S. soybean sales at $29.6 billion in 2008/2009 (USDA, 2010). The development of low phytate

soybeans has been a huge research interest due to the economical significance of the crop. Phytate is indigestible in non-ruminants, and since corn/soybean meal is common for livestock feed, low phytate soybeans are desirable in order to increase the available P to animals without the need to supplement with Pi or phytase (Wilcox et al., 2000). Several low phytate mutants have been identified and characterized in soybean. In the following section, each of the following low phytate mutants is discussed separately in detail: LR33, V99-5089, M766 and M153, and CX1834.

LR33

In order to induce mutations that would result in low phytate mutants, soybean seeds were soaked in *N*-nitroso-*N*-methylurea (Sebastian et al., 2000). The mutant identified, specifically "LR33," has about 50% lower phytate levels and also reduced levels of raffinosaccharides compared to the wild type. Raffinosaccharides consist of the raffinose and stachyose class of oligosaccharides and are also indigestible in monogastric animals. Animal feed made from crops with lower levels of these raffinosaccharides would be beneficial for animal nutrition (Hitz et al., 2002), similar to phytate reduction. *Myo*-inositol (a precursor to phytic acid) is known to be involved in the biosynthesis of raffinose and stachyose. Hitz et al. (2002) characterized the carbohydrates and other intermediates in the phytate and raffinosaccharide pathway of LR33 in order to gain insight into the actual gene causing the mutant phenotype. By measuring levels of *myo*-inositol, sucrose, galactinol, raffinose, and stachyose, the authors determined that the mutation in LR33 affects a step in the biochemical pathway that is needed for galactinol increase (Hitz et al., 2002). The authors also found that *myo*-inositol was decreased by 36% in LR33 compared to the wild type. These observations suggested decreased activity of MIPS1, and upon sequencing, it was determined that the MIPS1

gene from LR33 had a mutation of a G to T at nucleotide 1,188 of the coding sequence (Hitz et al., 2002). This causes an amino acid change from a lysine to a glutamine.

Gm-lpa-TW-1 and Gm-lpa-ZC-2

In order to develop low phytate mutants, seeds from the commercial soybean varieties ôTaiwan 75ô and ôZhechun no. 3ô were mutagenized by gamma irradiation (Yuan et al., 2007). $M_{2:3}$ seeds were harvested and screened for the high Pi/low phytate phenotype. Two low phytate mutants were identified: *Gm-lpa*-ZC-2 and *Gm-lpa*-TW-1, with a 46% and 66% reduction in phytate, respectively. Inheritance tests indicated that in both mutants, low phytate is under the control of a single, recessive gene. Molecular mapping was carried out to further delineate the genetic basis of low phytate in these two lines. The authors mapped 180 SSRs in an F_2 population and found Satt168 and Satt416 on LG B2 to be closely linked to the low phytate loci in *Gm-lpa*-ZC-2. For *Gm-lpa*-TW-1, they considered MIPS1 as a candidate gene because *Gm-lpa*-TW-1 displayed the same phenotype as mutant line LR33 (lowered phytate and raffinosaccharide levels). After sequencing the MIPS1 gene in *Gm-lpa*-TW-1 and comparing it to its parent, Taiwan 75, the authors found a 2 bp deletion in the third exon of this low phytate mutant. This most likely causes a loss of function in the protein due to the frame shift (Yuan et al., 2007).

V99-5089

There are three main forms of carbohydrates found in soybean seeds: sucrose, raffinose, and stachyose. As stated previously, raffinose and stachyose are known collectively as raffinosaccharides. The majority of seed carbohydrates is sucrose, with levels normally at 5-7%, followed by stachyose (3-4%), and raffinose (1%) (Skoneczka et al., 2009). The

raffinosaccharides are indigestible in monogastric animals and supply little metabolizable energy, however sucrose is beneficial because it is completely digestible by monogastric animals (Karr-Lilienthal et al., 2005). V99-5089 is a soybean line developed at Virginia Tech by conventional plant breeding, and is low phytate, low stachyose, and high sucrose. Due to these characteristics, V99-5089 has the potential for use in soybean breeding programs as a source of low phytate and raffinosaccharides. The low phytate/low stachyose/high sucrose QTL was mapped to LG B1, closely linked to Satt453, and the low phytate/raffinosaccharides phenotype in V99-5089 is known to be due to a mutation in MIPS1 (Biyashev et al., unpublished).

M766 and M153

Wilcox et al (2000) isolated two low phytate mutants in soybean: M153 and M766. These mutants were developed by soaking the seeds from the soybean breeding line ǿCX1515-4ô in EMS for 24 hours and then growing the M_1 plants. M_3 seeds were subsequently screened for the high Pi phenotype, which results from being homozygous for a low phytate mutation. The mutants were descendent from M_2 plants, and found to have nonlethal, heritable mutations similar to low phytate mutants previously reported in other crops. In M153 and M766, phytate in seeds was reduced by at least 50% in comparison to wild-type soybean seed. This reduction had no effect on total seed P because a reduction in phytate P was accompanied by an increase in Pi.

Gillman et al. (2009) investigated the genetic basis of low phytate in M766, and found a SNP mutation in an intron of the MRP ABC transporter gene on chromosome 3 (LG N), as well as a SNP mutation in the MRP gene on chromosome 19 (LG L). The mutation on chromosome 19 is at position 4864 relative to the start codon, is a nucleotide change from an A to T, and results in a stop codon. However, it is unclear if these mutations in M766 are the cause of low phytate, as it has not yet been evaluated in a segregating population.

CX1834

The following section discusses the key studies of CX1834, which is the main soybean low phytate line utilized in the research presented in the following chapters. This section covers the pertinent studies involving this line.

In 2000, Purdue University and the USDA-ARS, used the breeding line óCX1515-4ô (normal phytate levels), and treated it with EMS to induce mutations (Wilcox et al., 2000). Two M_2 low phytate progeny were identified: M766 (discussed previously) and M153. These are two different non-lethal mutants with heritable decreases in phytate with at least a 50% decrease in phytate P (Wilcox et al., 2000).

óAthow,ô a soybean breeding line with normal levels of phytate, was crossed with M153-1-4-6-14 in order to characterize the inheritance of low phytate mutant M153. The $F_{3:5}$ line CX1834-1-6 was selected as the low phytate progeny with the highest yield, although it was assumed that there was one allele controlling low phytate (Oltmans et al., 2004). At the Agricultural Engineering and Agronomy Research Center near Ames, IA, CX1834-1-6 was crossed with another normal phytate line, óA00-711013,ô to develop a segregating population. All F_1 progeny observed were normal phytate, indicating the low phytate alleles must be recessive. The F_2s were phenotyped, and out of 210, 197 had normal phytate levels, and 13 had low phytate levels. This indicated a 15:1 segregation ratio, which fits a model of two recessive loci with duplicate dominant epistasis (Oltmans et al., 2004). F_4 seeds derived from each of the 210 F_2 lines were: 86 normal phytate, 114 segregating, and 10 low phytate, which fits at 7:8:1 ratio, providing more evidence for a two gene model. These two low phytate loci in CX1834 were named *pha1* and *pha2* at the time (Oltmans et al., 2004).

In order to determine chromosomal locations of the low phytate genes in CX1834 (*pha1* and *pha2*), groups at the University of Tennessee and University of Georgia used three different mapping populations consisting of crosses between a normal phytate line and CX1834-1-2 (Walker et al., 2006). They determined that low phytate in CX1834 is a quantitative trait, and not a single gene, because if it was a single gene 25% of individuals would have low phytate (similar phytate levels as CX1834), but only 6.6% had low phytate. This suggested that low phytate is controlled by more than one locus (Walker et al., 2006). Single factor ANOVA, multiple regression analysis, interval mapping, and composite interval mapping showed a single locus on chromosome 3 (LG N) between markers Satt339 and Satt237 to be associated with phytate (R^2=40). A second locus was identified on chromosome 19 (LG L) between Satt527 and Satt561 (R^2=11) (Figure 1.3). Individuals homozygous for both mutated alleles had the lowest phytate levels. Loci on LGs L and N are most likely on duplicated regions of the genome, and the authors hypothesized that the two loci have a common origin and may encode related proteins. The authors also indicated that epistasis between the loci on L and N contributes to low phytate (R^2=0.08%). Individuals homozygous for the wild-type allele on LG N were all high phytate regardless of the allele on LG L. Homozygous and heterozygous individuals for the CX1834 allele on LG N were lower phytate than normal (midrange), IF they had the CX18343 allele for LG L. Individuals homozygous for both LG L and N loci had low phytate, with phytate levels in the range of the low phytate line CX1834. Heterozygosity at either locus gave mid-range phytate (Walker et al., 2006). These observations indicate incomplete dominance, as opposed to the duplicate dominant epistasis described by Oltmans et al. (2004), however it is unclear why there is a discrepancy and it may be due to differences in the Pi assay that each group used.

The authors also addressed the likelihood of obtaining two nonlethal mutations in two phytate-related genes. They noted that certain families of genes are more likely to be mutated than others (Koornneef et al., 1982), and if one mutation was already present in the line used for mutagenesis, it would likely be the locus located on L because the locus located on N has more of an effect on phenotype. There may be other minor genes affecting low phytate, that may go undetected due to the inability to identify polymorphic markers in all areas of the genome in order to complete a genome-wide scan (Walker et al., 2006). It was concluded that the markers Satt237 and Satt561 are diagnostic for low phytate genotypes and could be utilized for marker-assisted selection (MAS).

In general, there is a need to confirm discovered QTLs because mapping results may vary between different environments and with the use of different genotypes and phenotypic assays. In our lab (Gao et al., 2008), the results of Walker et al. (2006) were verified. A set of 208 $F_{2:5}$ lines of the population CX1834 x V99-3337 was used to verify the two low phytate QTLs determined previously by Walker et al. on LGs L and N. A QTL was detected on LG L near Satt527/Satt561 (R^2=12%), and a QTL was detected on LG N near Satt237/Satt339 (R^2=28%). The interaction between the two loci has an R^2 value of 12.9%. These results are similar to those reported by Walker et al (2006). The differences in R^2 values between the two independent studies may be due to different genetic backgrounds or different phytate assays (Gao et al., 2008). V99-3337, the other parent used in the cross, is a Virginia Tech experimental line with mid-range phytate. This mid-range phytate may be caused by unknown minor QTLs influencing low phytate. Also, phytate in our study was assayed directly following the procedure of Gao et al. (2007), rather than by measuring Pi and inferring the phytate levels indirectly. In addition, there may be other minor low phytate QTLs, suggested by the presence of high and low

transgressive segregants (offspring with significantly higher or lower phenotypic values than the parents). Five percent of the individuals had lower phytate than CX1834, and 70% had higher phytate than V99-3337 (Gao et al., 2008).

Scaboo et al. (2009) also confirmed the two QTL in different environments using two new populations, and confirmed that Satt237 and Satt561 are linked to the low phytate QTLs on LGs L and N. Individuals in the RIL populations that were homozygous for the CX1834 allele at both loci had the lowest phytate content (Scaboo et al., 2009). This was tested at three different locations with similar results. However, they found that using markers for MAS of the two loci was 50% less effective compared to phenotypic selection. That is, not all individuals with the CX1834 allele at Satt237 and Satt561 were actually low phytate. A heritability study of the low phytate allele showed that it was highly heritable; indicating that phytate in CX1834 is controlled mostly by genetics, with low influence from the environment. In addition, there were no differences in agronomic traits between low and normal phytate based on yield, height, lodging, maturity, protein, and oil content (Scaboo et al., 2009).

REFERENCES

Andaya C.B., Tai T.H. (2005) Fine mapping of the rice low phytic acid (*Lpa1*) locus. Theor. Appl. Genet. 111:489-95.

Anderson G., Arlidge E.Z. (1962) The adsorption of inositol phosphates and glycerophosphate by soil clays, clay minerals, and hydrated sesquioxides in acid media. Journal of Soil Science 13:216-224.

Antoine C., Lullien-Pellerin V., Abecassis J., Rouau X. (2004) Effect of wheat bran ball-milling on fragmentation and marker extractability of the aleurone layer. Journal of Cereal Science 40:275-282.

Bowen D.E., Souza E.J., Guttieri M.J., Raboy V., Fu J. (2007) A low phytic acid barley mutation alters seed gene expression. Crop Sci. 47:S-149-S-159.

Bregitzer P., Raboy V. (2006) Effects of four independent low-phytate mutations on barley agronomic performance. Crop Sci. 46:1318-1322.

Brinch-Pedersen H., Sørensen L.D., Holm P.B. (2002) Engineering crop plants: getting a handle on phosphate. Trends in Plant Science 7:118-125.

Campion B., Sparvoli F., Doria E., Tagliabue G., Galasso I., Fileppi M., Bollini R., Nielsen E. (2009) Isolation and characterisation of an *lpa* (low phytic acid) mutant in common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. 118:1211-1221.

Celi L., Lamacchia S., Marsan F.A., Barberis E. (1999) Interaction of inositol hexaphosphate on clays: adsorption and charging phenomena. Soil Science 164:574-585.

Corona M.E.P., Klundert I.v.d., Verhoeven J.T.A. (1996) Availability of organic and inorganic phosphorus compounds as phosphorus sources for *Carex* species. New Phytologist 133:225-231.

Cromwell G.L., Coffey R.D., Monegue H.J., Randolph J.H. (1995) Efficacy of low-activity, microbial phytase in improving the bioavailability of phosphorus in corn-soybean meal diets for pigs. J. of Anim. Sci. 73:449-457.

Dalai R.C. (1977) Soil Organic Phosphorus, in: N. C. Brady (Ed.), Advances in Agronomy, Academic Press. pp. 83-117.

Ertl D.S., Young K.A., Raboy V. (1998) Plant genetic approaches to phosphorus management in agricultural production. J Environ Qual 27:299-304.

Ficco D.B.M., Riefolo C., Nicastro G., De Simone V., Di Gesù A.M., Beleggia R., Platani C., Cattivelli L., De Vita P. (2009) Phytate and mineral elements concentration in a collection of Italian durum wheat cultivars. Field Crops Research 111:235-242.

Fileppi M., Galasso I., Tagliabue G., Daminati M., Campion B., Doria E., Sparvoli F. (2010) Characterisation of structural genes involved in phytic acid biosynthesis in common bean (*Phaseolus vulgaris* L.). Molecular Breeding 25:453-470.

Gao Y., Biyashev R.M., Saghai Maroof M.A., Glover N.M., Tucker D.M., Buss G.R. (2008) Validation of low phytate QTLs and evaluation of seedling emergence of low phytate soybeans. Crop Sci. 48:1355-1364.

Gao Y., Shang C., Saghai Maroof M.A., Biyashev R.M., Grabau E.A., Kwanyuen P., Burton J.W., Buss G.R. (2007) A modified colorimetric method for phytic acid analysis in soybean. Crop Sci. 47:1797-1803.

Gillman J.D., Pantalone V.R., Bilyeu K. (2009) The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. The Plant Genome 2:179-190.

Guttieri M., Bowen D., Dorsch J.A., Raboy V., Souza E. (2004) Identification and characterization of a low phytic acid wheat. Crop Sci. 44:418-424.

Guttieri M.J., Peterson K.M., Souza E.J. (2006) Agronomic performance of low phytic acid wheat. Crop Sci. 46:2623-2629.

Harrison A.F. (1987) Soil organic phosphorus. A review of world literature. CAB International., Oxford.

Haygarth P.M., Chapman P.J., Jarvis S.C., Smith R.V. (1998) Phosphorus budgets for two contrasting grassland farming systems in the UK. Soil Use and Management 14:160-167.

Hitz W.D., Carlson T.J., Kerr P.S., Sebastian S.A. (2002) Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiol. 128:650-660.

Holm P.B., Kristiansen K.N., Pedersen H.B. (2002) Transgenic approaches in commonly consumed cereals to improve iron and zinc content and bioavailability. The Journal of Nutrition 132:514S-516S.

Hurrell R.F. (2004) Phytic acid degradation as a means of improving iron absorption. Int J Vitam Nutr Res. 74:445-52.

Jang D., Fadel J., Klasing K., Mireles A., Jr, Ernst R., Young K., Cook A., Raboy V. (2003) Evaluation of low-phytate corn and barley on broiler chick performance. Poult Sci 82:1914-1924.

Karr-Lilienthal L.K., Kadzere C.T., Grieshop C.M., Fahey J.G.C. (2005) Chemical and nutritional properties of soybean carbohydrates as related to nonruminants: A review. Livestock Production Science 97:1-12.

Khan A.J., Akhtar, A., Farooq-I-Azam, Zeb, A. (2007) Identification and isolation of low phytic acid wheat (*Triticum Aestivum* L.) inbred lines/mutants. Pak. J. Bot. 39:2051-2058.

Kim S.-I., Tai T. (2010) Genetic analysis of two *OsLpa*1-like genes in *Arabidopsis* reveals that only one is required for wild-type seed phytic acid levels. Planta 232:1241-1250.

Kim S., Andaya C., Newman J., Goyal S., Tai T. (2008) Isolation and characterization of a low phytic acid rice mutant reveals a mutation in the rice orthologue of maize MIK. Theor. Appl. Genet. 117:1291-1301.

Koornneef M., Dellaert L.W.M., van der Veen J.H. (1982) EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Mutat. Res. 93:109-123.

Kroehler C.J., Linkins A.E. (1991) The absorption of inorganic phosphate from ^{32}P-labeled inositol hexaphosphate by *Eriophorum vaginatum*. Oecologia 85:424-428.

Larson S.R., Raboy V. (1999) Linkage mapping of maize and barley *myo*-inositol 1-phosphate synthase DNA sequences: correspondence with a low phytic acid mutation. Theor. Appl. Genet. 99:27-36.

Larson S.R., Young K.A., Cook A., Blake T.K., Raboy V. (1998) Linkage mapping of two mutations that reduce phytic acid content of barley grain. Theor. Appl. Genet. 97:141-146.

Lei X., Ku P.K., Miller E.R., Ullrey D.E., Yokoyama M.T. (1993) Supplemental microbial phytase improves bioavailability of dietary zinc to weanling pigs. The Journal of Nutrition 123:1117-1123.

Lemtiri-Chlieh F., MacRobbie E.A.C., Webb A.A.R., Manison N.F., Brownlee C., Skepper J.N., Chen J., Prestwich G.D., Brearley C.A. (2003) Inositol hexakisphosphate mobilizes an endomembrane store of calcium in guard cells. PNAS 100:10091-10095.

Li M., Osaki M., Madhusudana Rao I., Tadano T. (1997) Secretion of phytase from the roots of several plant species under phosphorus-deficient conditions. Plant and Soil 195:161-169.

Liu Q.-L., Xu X.-H., Ren X.-L., Fu H.-W., Wu D.-X., Shu Q.-Y. (2007) Generation and characterization of low phytic acid germplasm in rice (*Oryza sativa* L.). Theor. Appl. Genet. 114:803(12).

Loewus F.A., Murthy P.P.N. (2000) *Myo*-inositol metabolism in plants. Plant Science 150:1-19.

Lott J.N.A. (1980) Protein Bodies, in: N. E. Tolbert (Ed.), The Biochemistry of Plants, Academic Press, New York. pp. p.589-623.

Lott J.N.A., Greenwood J.S., Batten G.D. (1995) Mechanisms and regulation of mineral nutrient storage during seed development, in: G. G. Kigel J (Ed.), Seed Development and Germination., Marcel Dekker, New York. pp. 215ï 235.

Morrison I.N., Kuo J., O'Brien T.P. (1975) Histochemistry and fine structure of developing wheat aleurone cells. Planta 123:105-116.

Morse D., Head H.H., Wilcox C.J. (1992) Disappearance of phosphorus in phytate from concentrates in vitro and from rations fed to lactating dairy cows. Journal of Dairy Science 75:1979-1986.

Nagy R., Grob H., Weder B., Green P., Klein M., Frelet-Barrand A., Schjoerring J.K., Brearley C., Martinoia E. (2009) The *Arabidopsis* ATP-binding cassette protein AtMRP5/AtABCC5 is a high affinity inositol hexa*kis*phosphate transporter involved in guard cell signaling and phytate storage. Journal of Biological Chemistry 284:33614-33622.

O'Dell B.L., De Boland A.R., Koirtyohann S.R. (1972) Distribution of phytate and nutritionally important elements among the morphological components of cereal grains. Journal of Agricultural and Food Chemistry 20:718-723.

Oliver R.E., Yang, C., Hu, G., Raboy, V., Zhang, M. (2009) Identification of PCR-based DNA markers flanking three low phytic acid mutant loci in barley. Journal of Plant Breeding and Crop Science 1:87-93.

Oltmans S.E., Fehr W.R., Welke G.A., Cianzio S.R. (2004) Inheritance of low-phytate phosphorus in soybean. Crop Sci. 44:433-435.

Panzeri D., Cassani E., Doria E., Tagliabue G., Forti L., Campion B., Bollini R., Brearley C.A., Pilu R., Nielsen E., Sparvoli F. (2011) A defective ABC transporter of the MRP family, responsible for the bean *lpa1* mutation, affects the regulation of the phytic acid pathway, reduces seed *myo*-inositol and alters ABA sensitivity. New Phytologist:70-83.

Pilu R., Panzeri D., Gavazzi G., Rasmussen S.K., Consonni G., Nielsen E. (2003) Phenotypic, genetic and molecular characterization of a maize low phytic acid mutant (*lpa241*). Theor. Appl. Genet. 107:980-7.

Planckaert F., Walbot V. (1989) Molecular and genetic characterization of *Mu* transposable elements in *Zea mays*: behavior in callus culture and regenerated plants. Genetics 123:567-78.

Poulsen H.D. (2000) Phosphorus utilization and excretion in pig production. Journal of Environmental Quality 29:24-27.

Raboy V. (2001) Seeds for a better future: 'low phytate' grains help to overcome malnutrition and reduce pollution. Trends Plant Sci. 6:458-462.

Raboy V. (2007) Seed phosphorus and the development of low-phytate crops, in: B. L. Turner, Richardson, A.E., and Mullaney, E.J. (Ed.), Inositol Phosphates: Linking Agriculture and the Environment, CAB International, Oxfordshire, UK. pp. 111-132.

Raboy V. (2009) Seed Total Phosphate and Phytic Acid, Molecular Genetic Approaches to Maize Improvement. pp. 41-53.

Raboy V., Gerbasi P. (1996) Genetics of *myo*-inositol phosphate synthesis and accumulation. Subcell Biochem 26:257-285.

Raboy V., Gerbasi P.F., Young K.A., Stoneberg S.D., Pickett S.G., Bauman A.T., Murthy P.P.N., Sheridan W.F., Ertl D.S. (2000) Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1. Plant Physiol. 124:355-368.

Rasmussen S.K., Ingvardsen C.R., Torp A.M. (2010) Mutations in genes controlling the biosynthesis and accumulation of inositol phosphates in seeds. Biochemical Society Transactions 38:689-694.

Scaboo A.M., Pantalone V.R., Walker D.R., Boerma H.R., West D.R., Walker F.R., Sams C.E. (2009) Confirmation of molecular markers and agronomic traits associated with seed phytate content in two soybean RIL populations. Crop Sci. 49:426-432.

Sebastian S.A., Kerr P.S., Pearlstein R.W., Hitz W.D. (2000) Soybean germplasm with novel genes for improved digestibility, in: J. K. Drackley (Ed.), Soy in Animal Nutrition, Federation of Animal Science Societies, Savoy, IL. pp. 56-74.

Shi J., Wang H., Hazebroek J., Ertl D.S., Harp T. (2005) The maize low-phytic acid 3 encodes a *myo*-inositol kinase that plays a role in phytic acid biosynthesis in developing seeds. The Plant Journal 42:708-719.

Shi J., Wang H., Wu Y., Hazebroek J., Meeley R.B., Ertl D.S. (2003) The maize low-phytic acid mutant *lpa2* is caused by mutation in an inositol phosphate kinase gene. Plant Physiol. 131:507-515.

Shi J., Wang H., Schellin K., Li B., Faller M., Stoop J.M., Meeley R.B., Ertl D.S., Ranch J.P., Glassman K. (2007) Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. Nat. Biotech 25:930-937.

Shukla S., VanToai T.T., Pratt R.C. (2004) Expression and nucleotide sequence of an INS (3) P1 synthase gene associated with low-phytate kernels in maize (*Zea mays* L.). J. Agric. Food Chem. 52:4565-4570.

Skoneczka J.A., Saghai Maroof M.A., Shang C., Buss G.R. (2009) Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. Crop Sci. 49:247-255.

Stevenson-Paulik J., Bastidas R.J., Chiou S.-T., Frye R.A., York J.D. (2005) Generation of phytate-free seeds in *Arabidopsis* through disruption of inositol polyphosphate kinases. PNAS 102:12612-12617.

Turner B.L., Papházy M.J., Haygarth P.M., Mckelvie I.D. (2002) Inositol phosphates in the environment. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357:449-469.

USDA. (2010) Soybeans and Oil Crops.

Veum T.L., Ledoux D.R., Bollinger D.W., Raboy V., Cook A. (2002) Low-phytic acid barley improves calcium and phosphorus utilization and growth performance in growing pigs. J. Anim Sci. 80:2663-2670.

Walker D.R., Scaboo A.M., Pantalone V.R., Wilcox J.R., Boerma H.R. (2006) Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. Crop Sci. 46:390-397.

Wilcox J.R., Premachandra G.S., Young K.A., Raboy V. (2000) Isolation of high seed inorganic P, low-phytate soybean mutants. Crop Sci. 40:1601-1605.

Xu X.-H., Zhao H.-J., Liu Q.-L., Frank T., Engel K.-H., An G., Shu Q.-Y. (2009) Mutations of the multi-drug resistance-associated protein ABC transporter gene 5 result in reduction of phytic acid in rice seeds. Theor. Appl. Genet. 119:75-83.

Yuan F.-J., Zhao H.-J., Ren X.-L., Zhu S.-L., Fu X.-J., Shu Q.-Y. (2007) Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). Theor. Appl. Genet. 115:945-957.

Zhao J., Jamar D.C.L., Lou P., Wang Y., Wu J., Wang X., Bonnema A.B., Koornneef M., Vreugdenhil D. (2008) Quantitative trait loci analysis of phytate and phosphate concentrations in seeds and leaves of *Brassica rapa*.

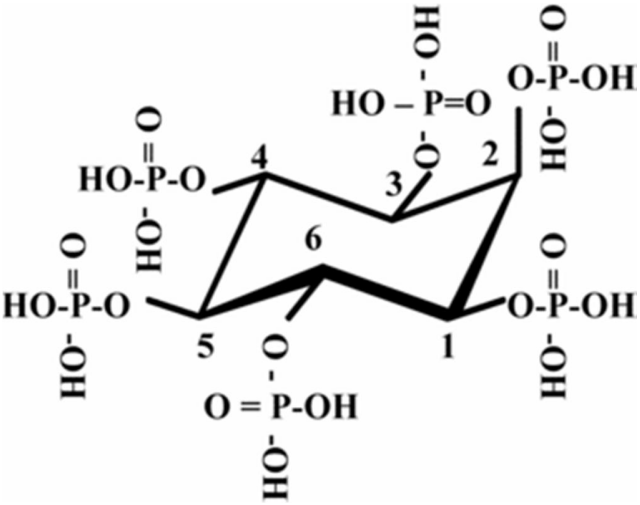


Figure 1.1. Chemical structure of phytic acid, shown in the boat formation with five equatorial and one axial phosphate groups. [Fair use] Rasmussen S.K., Ingvardsen C.R., Torp A.M. (2010) Mutations in genes controlling the biosynthesis and accumulation of inositol phosphates in seeds. Biochemical Society Transactions 38:689-694. http://www.biochemsoctrans.org/bst/038/0689/bst0380689.htm (accessed July 10, 2011). Fair use determination attached.

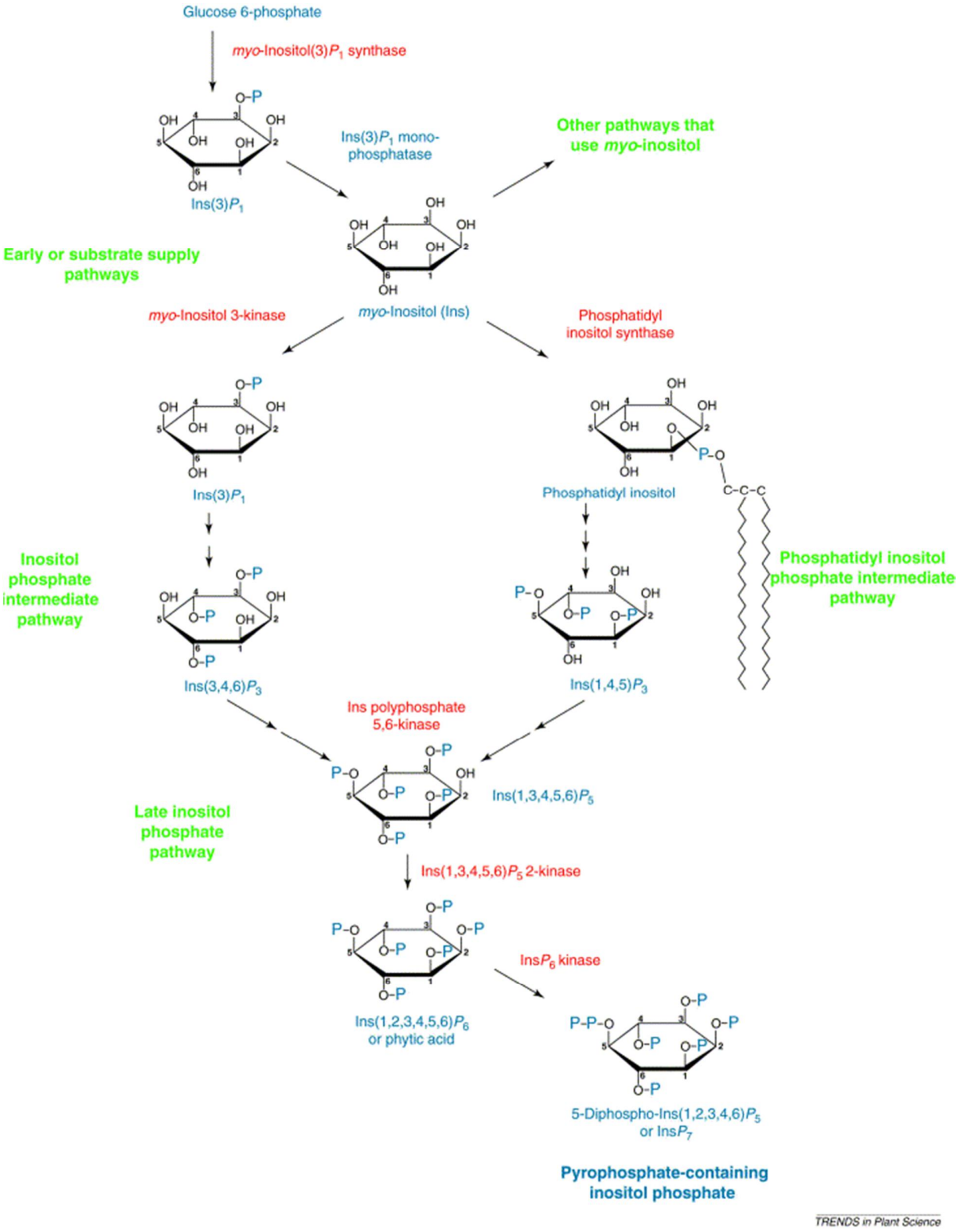

Glucose 6-phosphate
myo-Inositol(3)P1 synthase
Ins(3)P1
Ins(3)P1 mono-phosphatase
Other pathways that use myo-inositol
Early or substrate supply pathways
myo-Inositol (Ins)
myo-Inositol 3-kinase
Phosphatidyl inositol synthase
Ins(3)P1
Phosphatidyl inositol
Inositol phosphate intermediate pathway
Phosphatidyl inositol phosphate intermediate pathway
Ins(3,4,6)P3
Ins(1,4,5)P3
Ins polyphosphate 5,6-kinase
Ins(1,3,4,5,6)P5
Late inositol phosphate pathway
Ins(1,3,4,5,6)P5 2-kinase
InsP6 kinase
Ins(1,2,3,4,5,6)P6 or phytic acid
5-Diphospho-Ins(1,2,3,4,6)P5 or InsP7
Pyrophosphate-containing inositol phosphate
TRENDS in Plant Science

Figure 1.2. Biosynthesis of phytic acid in a eukaryotic cell. [Fair use] Raboy V. (2001) Seeds for a better future: 'low phytate' grains help to overcome malnutrition and reduce pollution. Trends Plant Sci. 6:458-462. http://www.sciencedirect.com/science/article/pii/S1360138501021045 (accessed July 10, 2011). Fair use determination attached.

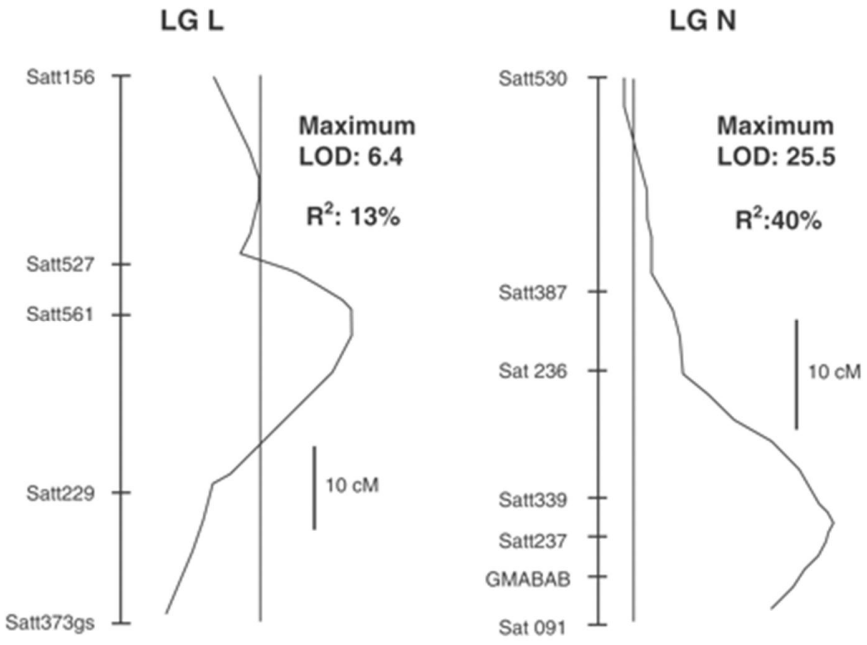


Figure 1.3. QTL maps of low phytate loci in CX1834. Low phytate was shown to be associated with loci on LGs L and N. [Fair use] Walker D.R., Scaboo A.M., Pantalone V.R., Wilcox J.R., Boerma H.R. (2006) Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. Crop Sci. 46:390-397. https://www.agronomy.org/publications/cs/articles/46/1/390 (accessed July 10, 2011). Fair use determination attached.

II. THE GENETIC BASIS OF THE LOW PHYTATE TRAIT IN THE SOYBEAN LINE CX1834

M. A. Saghai Maroof[a,*], Natasha M. Glover[a], Ruslan M. Biyashev[a], Glenn R. Buss[a], Elizabeth A. Grabau[b]

[a] Department of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA, 24061

[b] Department of Plant Pathology, Physiology and Weed Science, Virginia Tech, Blacksburg, VA, 24061

Received_____* Corresponding author (smaroof@vt.edu)

Key Words: soybean, low phytate, *myo*-inositol phosphate synthase, MIPS, multidrug resistance-associated protein, MRP, ABC transporter, CX1834

Abbreviations: ABC, ATP-binding cassette; LG, linkage group; LOD, logarithm of the odds; LP, low phytate; *lpa*, low phytic acid; MAS, marker-assisted selection; MIPS, *myo*-inositol phosphate synthase; MRP, multidrug resistance-associated protein; PCR, polymerase chain reaction; QTL, quantitative trait loci; RIL, recombinant inbred line; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; WGS, whole genome shotgun

ABSTRACT

The low phytate trait in plant seeds offers important nutritional and environmental benefits for food and feed uses. Mutants with reduced phytate content are commonly produced by chemical mutagenesis, as is the case for soybean [*Glycine max* (L.) Merr.] line CX1834. Quantitative trait loci (QTL) for low phytate in CX1834 have previously been reported, however the genetic basis for this trait has not been identified. In this study, we examined several possible chromosomal map locations for the low phytate mutation in CX1834. After eliminating the *myo*-inositol phosphate synthase (MIPS) gene family as the location of the low phytate mutations based on mapping studies, we focused on candidate genes in regions of known low phytate QTL on linkage groups (LGs) L and N. Using the soybean whole genome sequence, we identified genes encoding two putative multidrug resistance-associated proteins (MRPs). The sequences from segments of the CX1834 MRP genes on LGs L and N were each compared with sequences from 16 other soybean lines with normal phytate content. In CX1834, a single nucleotide mutation from A to T, resulting in substitution of a stop codon for an arginine residue, was detected in the MRP gene on LG N of the low phytate line, but not in the normal counterparts. In addition, a single nucleotide mutation from a G to A was detected in CX1834 but not the wild type in the LG L MRP gene. Further comparative sequence analysis of normal and low phytate progeny from a cross of CX1834 with V99-3337 (normal phytate) also indicated the A to T substitution on the LG N MRP, and the G to A substitution on the LG L MRP in low phytate individuals, suggesting mutations in MRP genes as the possible cause of the low phytate phenotype.

INTRODUCTION

Phytic acid, otherwise known as *myo*-inositol (1,2,3,4,5,6)-hexa*kis*phosphate, is the major storage form of phosphorus in soybean [*Glycine max* (L.) Merr.] and other plant seeds. Phytate (the mixed salt form of phytic acid found in seeds) typically comprises 75% of total plant seed phosphorus (Raboy, 2001). The poor digestibility of phytate and its chelation of mineral cations contribute to its classification as an antinutrient (Lott et al., 1995). Non-ruminant livestock such as poultry and swine excrete unavailable phytate, which necessitates supplementation with inorganic phosphorus to meet phosphorus growth requirements. Another solution is to include phytase enzyme supplements in feed, which results in release of up to 50% of the phosphorus from the seed. This reduces the environmental consequences, but contributes significantly to feed costs (Cromwell et al., 1995; Raboy, 2001). Thus, low phytate soybeans are desirable from both a nutritional and environmental standpoint and also provide an economic advantage to producers.

Few, if any, naturally occurring low phytate crops have been identified. However, low phytate mutants have been isolated in a number of crops, including barley, (Larson et al., 1998), rice (Larson et al., 2000), wheat (Guttieri et al., 2004), maize (Raboy et al., 2000), and soybean (Wilcox et al., 2000) as a result of chemical mutagenesis or gamma irradiation. Mutations are introduced into one or more genes, thereby disrupting or decreasing their function in phytic acid biosynthesis or accumulation (Raboy, 2007). In several cases, the genes responsible for the low phytate phenotype have been identified, including MIPS (Hitz et al., 2002), *myo*-inositol kinases (Shi et al., 2005), inositol polyphosphate kinases (Shi et al., 2003), and multidrug resistance-associated protein ATP-binding cassette (MRP ABC) transporters (Shi et al., 2007).

Myo-inositol phosphate synthase, a key enzyme catalyzing the first step in phytic acid biosynthesis, is encoded by four genes in soybean, MIPS1 to MIPS4 (Chappell et al., 2006; Hegeman et al., 2001). In two different low phytate soybean lines, the low phytate trait was shown to be associated with independent mutations in the MIPS1 gene. More specifically, the mutation in low phytate line LR33 resulted from a substitution of G to T in exon 10 (Hitz et al., 2002), and the low phytate *Gm-lpa*-TW-1 mutant involved a two-nucleotide deletion in exon 3 of MIPS1 (Yuan et al., 2007). In a third low phytate soybean line, *Gm-lpa*-ZC-2, the low phytate QTL mapped to LG B2 but the mutated gene is unknown (Yuan et al., 2007).

A fourth low phytate soybean line, CX1834, is derived from a cross between óAthowô and M153-1-4-6-14, where Athow is a soybean cultivar with normal phytate content and M153-1-4-6-14 is a low phytate mutant developed at USDA/ARS-Purdue University (Wilcox et al., 2000). Low phytate in CX1834 is controlled by two recessive genes (Oltmans et al., 2004), and two QTL controlling low phytate in CX1834 were mapped onto LGs L and N by Walker et al. (2006) and were validated in a different background by Gao et al. (2008). However, it is unknown what underlying genes in these QTL regions control the low phytate phenotype. To address this question, two approaches were undertaken to elucidate the low phytate trait in CX1834. The first objective included mapping all four MIPS genes in soybean to examine their correspondence with the known low phytate QTLs in CX1834, and subsequently, the second objective involved identifying candidate genes in CX1834 using the soybean whole genome shotgun sequence (WGS), which only recently has become available at www.phytozome.net/soybean.php (Soybean Genome Project, DOE Joint Genome Institute).

MATERIALS AND METHODS

Genetic Materials

Three low phytate lines and fifteen normal phytate lines were used in this study. Low phytate lines included CX1834-1-6 (hereafter CX1834), M766 (Wilcox et al., 2000) and V99-5089 (Gao et al., 2008). Normal phytate lines tested included Williams 82, óEssex,ô óLee 68,ô óHutcheson,ô óAthow,ô óWye,ô óMN1401,ô R95-1705, V99-8060, V99-3337, V71-370, PI 200508, PI 96983, PI 407162, and PI 87013 (Table 2.1).

Two different populations were used for mapping: 1) 208 F_2 individuals from CX1834 x V99-3337, where V99-3337 is a Virginia Tech experimental line developed from Hutcheson (Buss et al., 1988), and 2) 301 F_8 RILs from the interspecific cross between V71-370 x PI 407162 (Maughan et al., 2000). V71-370 is a *G. max* breeding line and PI 407162 is a *G. soja* plant introduction. Ten progeny lines were selected from the $F_{2:5}$ generation of the CX1834 x V99-3337 population to examine phytate content and candidate gene status. The ten lines, which included five high phytate (designated as H1 to H5) and five low phytate (designated as L1 to L5) selections (Table 2.1) were used for gene sequencing. In addition, from the same cross, one BC_1F_2 plant designated as ó03-516-4ô was selected which contained homozygous CX1834 alleles at known marker loci linked to the low phytate trait. Two BC_1F_4 lines, 04-1272 and 04-897, were also included in this study. The two lines were developed from BC_1F_2 03-516-4 and contained phytate levels even lower than the donor low phytate line, CX1834 (Table 2.1, see also Gao et al., 2008). All phytate assays were performed according to the colorimetric method of Gao et al. (2007).

Molecular Marker Analysis

To map MIPS genes, primers were designed from MIPS1, MIPS2, MIPS3, and MIPS4 genomic sequences (GenBank accession number DQ323904, DQ323906, DQ323907, and DQ323908, respectively). A total of 26 primer pairs were designed to identify polymorphisms in the four MIPS genes: One primer pair for MIPS1, two for MIPS2, fifteen for MIPS3, and eight for MIPS4. The MIPS1 primer pair was designed to detect alleles with or without a known 39 bp deletion polymorphism as described by Chappell et al. (2006). MIPS2, MIPS3 and MIPS4 primers targeted amplification of simple sequence repeats (SSRs). MIPS1 through MIPS4 primers were screened for polymorphism among 15 soybean lines (Table 2.1).

A total of 12 primer pairs, also targeting SSRs, were designed to test for polymorphism in two MRP genes from the soybean WGS. The putative soybean MRP genes were identified in the soybean genome utilizing the maize MRP4 gene sequence (GenBank Accession number EF586878) and soybean MRP partial DNA sequence GM-MRP (Shi et al., 2007) as query sequences in a Blast search against the WGS. Primers used for mapping MRPs were screened for polymorphism using eight soybean lines (Table 2.1).

Genomic template sequences were examined for SSR presence (except MIPS1) using the computer software SSRIT (Temnykh et al., 2001), available online at http://www.gramene.org/db/searches/ssrtool. After target SSR motifs were identified, appropriate primers were designed online at http://frodo.wi.mit.edu/ using the computer program Primer 3 (Rozen and Skaletsky, 1999).

DNA Extraction and Gel Electrophoresis

Leaf tissue samples from individual F_2 plants and bulked samples from each of five low phytate (L1 to L5), five high phytate (H1 to H5) selections (Gao et al., 2008) and parental lines

were collected for DNA extraction. Genotyping procedures and marker data collection including DNA isolation, polymerase chain reaction, gel electrophoresis and SSR assays were performed with radioactive polyacrylamide gel electrophoresis according to Yu et al. (1994) and Saghai Maroof et al. (1994), or with M13 dye labeled primers using an Applied Biosystems 3130xl Genetic Analyzer (Foster City, CA). Genetic Analyzer-generated data were analyzed using the genotyping software GeneMarker version 1.60 designed by SoftGenetics (State College, PA, USA).

Map Construction

JoinMap 4 (Van Ooijen, 2006) was used for linkage group construction for mapping MIPS genes at an LOD of 3.0 based on the Kosambi function. Using polymorphisms from MIPS1 and MIPS2, the genes were mapped in the F_2 population of CX1834 x V99-3337 using 94 individuals. MIPS3 was mapped in the F_8 population of V71-370 x PI 407162 using 301 individual RILs. Mapping the MIPS4 gene through this strategy was not possible due to the lack of polymorphism within the gene. An additional approach for mapping MIPS4 gene was employed based on the soybean whole genome sequence (see results section). For this purpose, SSRs from the flanking regions of the MIPS4 gene were targeted and primers were designed for mapping. An SSR marker for the MIPS4 gene, designated SF22MIPS4, was mapped in V71-370 x PI 407162 RIL population.

QTL Mapping

QTL data and maps were adapted from Gao et al. (2008). To generate LOD curves, Mapmaker 3.0b (Lander et al., 1987) was used for linkage group construction at an LOD of 3.0 using the ñgroupò command with a maximum Haldane distance of 50 cM. Putative QTLs were

located by interval mapping using QTL Cartographer 2.5 (Wang et al., 2007) with an LOD threshold established at 4.8 by l000 permutation tests (Churchill and Doerge, 1994).

DNA Sequencing

Polymerase chain reaction (PCR)-amplified products were used for DNA sequencing. PCR procedures were as follows: 1X PCR buffer, 0.2 mM of each dNTP, 1.5 mM of $MgCl_2$, each primer 0.5 uM, 50 ng of genomic DNA template, 2 units of Taq DNA Polymerase, and PCR grade water to final volume of 50 ul per reaction. AmpliTaq Gold (Taq DNA Polymerase) manufactured by Roche (Branchburg, New Jersey, USA) was used for all PCR amplifications designated for subsequent sequencing. After amplification, all PCR products were verified for size and integrity by electrophoresis on 1% agarose gels in TAE buffer following standard procedures (Sambrook, 1989). PCR products were purified using QIAquick PCR Purification Kit (QIAGEN Sciences, Maryland, USA) according to manufacturerôs protocol. In some cases when more than one PCR product was amplified, the expected size band was excised from the agarose gel and DNA was extracted using QIAEX II kit (QIAGEN Sciences, Maryland, USA) according to manufacturerôs protocol and then purified with QIAquick PCR Purification Kit. After purification, all DNA templates were quantified using an ND-1000 NanoDrop spectrophotometer (NanoDrop Technologies LLC, Wilmington, Delaware, USA) following manufacturerôs protocols and then submitted to the Core Laboratory Facility of Virginia Bioinformatics Institute at Virginia Tech for sequencing by automated capillary electrophoresis. All PCR-generated DNA templates were sequenced in both directions. Forward and reverse sequences were assembled into a contig using SeqMan, trimmed using EditSeq, and aligned with the MegAlign through the Clustal W method. All these DNA sequence analysis programs are

part of the DNASTAR Lasergene 7.1 software package (DNASTAR, Inc., Madison, Wisconsin, USA).

RESULTS

Mapping MIPS Genes

MIPS genes were mapped in order to determine whether they had any genetic association with the known low phytate QTLs in CX1834 located on LGs L and N. A total of 26 primer pairs were designed to amplify polymorphism from MIPS sequences for mapping. All primers designed to map MIPS 1-4 genes were used to screen for polymorphisms in a set of soybean lines (Table 2.1). The three pairs of primers used for mapping MIPS 1 to 3 are shown in Table 2.2.

Mapping data revealed that MIPS1 is located on LG B1, MIPS2 is located on one end of LG G, and MIPS3 is located on LG A1 (Figure 2.1). As mentioned previously, it was not possible to map MIPS4 by this method due to lack of polymorphism. Therefore, a different strategy based on the use of the soybean WGS was employed for marker development in order to map MIPS4. Full genomic sequence of MIPS4 gene was used in a BLAST homology search against the soybean WGS. Sequence alignment allowed us to narrow down the number of critical hits for the MIPS4 gene to one supercontig/scaffold. At the time of this study, the WGS was available only as a draft, in a fragmented format. Scaffolds existed as supercontigs and their associations with LGs had not been established yet. To map the MIPS4 gene, sequences from the scaffold region adjacent to the MIPS4 gene were downloaded, scanned for SSR presence, and used to design primers. Using the primers, eight soybean lines (Table 2.1) were screened for polymorphism. Primer SF22MIPS4 (Table 2.2) detected polymorphism between the parental lines V71-370 and PI407-182. Mapping results located MIPS4 onto LG A2 based on data from 141 RILs (Figure 2.1).

Mapping MRP Genes

A class of maize low phytate mutants, *lpa1*, is due to a defective MRP ABC transporter (Shi et al. 2007). To test the hypothesis that low phytate trait in CX1834 is due to a mutation in an MRP gene, the maize MRP4 gene and also the 555 bp MRP sequence (GM-MRP) of the soybean gene construct used to silence the soybean homolog of maize MRP4 by Shi et al. (2007), were used in a BLAST homology search against the WGS. Positive hits containing the most significant alignments were from scaffolds 63, 165 and 84 (www.phytozome.net/soybean.php). By superimposing genetic map positions of known single nucleotide polymorphism (SNP) and SSR markers with known NCBI annotations on the soybean WGS, we were able to determine that soybean MRP-related sequences are located on LG N (scaffold 63), LG L (scaffold 165), and LG F (scaffold 84). The MRPs located on LGs L and N were in the region of the low phytate QTLs based on prior mapping work (Walker et al. 2006; Gao et al. 2008). The MRP located on LG F was tested for association with the phytate trait by designing several sets of primers from within or around the MRP, and using the polymorphic markers to genotype the CX1834 x V99-3337 population. One-way ANOVA determined there was no significant association with phytate and any of the markers test (results not shown), so this MRP was ruled out as a candidate gene.

Because the L and N MRP genes are located in the same chromosomal regions as known CX1834 low phytate QTLs, they were considered candidate genes that may harbor a mutation conferring the low phytate phenotype and/or provide a marker for the trait. The first MRP-related sequence, designated SF63MRP, was located at scaffold_63:3359322..3369774 (+ strand) and is 10,453 bp (full genomic sequence), with a predicted protein length of 1389 amino acids originating from 16 exons (www.phytozome.net/soybean.php). The second MRP-related

sequence, designated SF165MRP, was located on scaffold_165:196992..204393 (+ strand), is 7,402 bp (full genomic sequence), with a predicted protein of 405 amino acids originating from ten exons.

In order to confirm that SF63MRP and SF165MRP are located on LGs N and L respectively, the MRPs were mapped. Primers were designed to target SSRs in these genes, and the primers used to map SF63MRP were designed from the third intron of the predicted gene. Primers from the introns of SF165MRP did not detect polymorphism, so primers were designed from the geneôs flanking region, 277 bp upstream from the start of the SF165MRP gene. SSR markers for both genes were identified and mapped in F_2 of CX1834 x V99-3337 using 94 individuals (Figure 2.2). SF63MRP mapped 2.6 cM from Satt237, and SF165MRP mapped 2.2 cM from Satt527. Both Satt237 (LG N) and Satt527 (LG L) have previously been reported to be tightly linked to low phytate QTLs in CX1834 (Walker et al., 2006; Gao et al., 2008). Because both MRP-related sequences map to the same regions as the low phytate QTLs, this provides further evidence that the low phytate trait in CX1834 is caused by a mutation in one or both of the identified candidate MRP sequences.

MRP Sequencing

Primers were designed to amplify coding sequences (exons) of the putative MRP genes located on scaffolds 63 and 165. Fourteen primer pairs were designed to amplify the 16 exons of SF63MRP, and ten primer pairs were designed to amplify the ten exons of SF165MRP (Table 2.3). Coding sequences of the MRP candidate genes were completely sequenced in CX1834. Exon sequences were aligned with the reference Williams 82 sequence from the WGS in order to search for mutations.

Sequence alignment of SF63MRP from CX1834, compared to Williams 82 from the WGS revealed a single base mutation (Figure 2.3A) at position 2368 of the complete coding sequence (equivalent to nucleotide position 45 of exon 9). This single base substitution from A (wild type) to T (mutant) in CX1834ôs SF63MRP gene results in a stop codon (Figure 2.3B), resulting in a predicted truncated protein product. Upon sequencing SF165MRP, one mutation was identified by comparing the sequence from CX1834 to the Williams 82 reference sequence and to the three other normal phytate lines: M766, V99-3337, and óAthow.ô The mutation is located in the predicted exon six, and a point mutation at position 3116 of the coding sequence: the wild-type is a G nucleotide, whereas the CX1834 type is an A (Fig 2.3C). This mutation causes a change from an arginine to a lysine (Fig 2.3D).

ABC transporter proteins have several domains that are needed to carry out their function, including a nucleotide-binding domain and a transmembrane domain, among others. In order to determine if the mutations detected were located in any domains in the protein which may affect their function, we used InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) to scan for signature domains in the LG L and N MRP genes. We found that neither of these mutations was located in any ABC transporter family signature domain.

Additional sequencing was done in order to further examine the association of the mutation observed in CX1834 with the low phytate trait. Exon nine from SF63MRP and exon six from SF165MRP were sequenced from a set of 16 soybean lines (Table 2.1), and two low phytate individual and bulk selections from CX1834 x V99-3337 cross (i.e., lines 1272, and 897 in Table 2.1). In addition, the five highest- and five lowest-phytate lines were selected from $F_{2:5}$ generation of CX1834-1-6 x V99-3337 (Gao et al., 2008) and sequenced. For SF63MRP, the exon nine sequences from the different lines was aligned to Williams 82 from the WGS and the

resulting sequence information clearly showed that the single nucleotide change of A to T, as described above, is the signature mutation for the low phytate line CX1834. All low phytate selections used in this study contained a T at position 2368 of the coding sequence, while all normal lines contained an A at the corresponding position. For SF165MRP, a similar result was seen, and in all cases, the five low phytate selections had an A at the mutation site and the five high phytate selections had either an A or G at the mutation site (Table 2.1). If the high phytate selection had an A mutation on SF165MRP, it did not have the LG N mutation. This is in agreement with previous research stating that both low phytate alleles are needed to confer low phytate (Gillman et al., 2009; Oltmans et al., 2004; Walker et al., 2006). Therefore, only individuals with the two mutations combined were low phytate, one mutation at either L or N mutation site is not enough to confer low phytate.

DISCUSSION

The availability of the draft soybean genome sequence (Soybean Genome Project, DOE Joint Genome Institute) provides a powerful tool for genomics-related studies including putative gene identification, targeted sequence amplification, marker development and verification of mapping results (Jackson et al., 2006). In this study, the WGS resources were efficiently used to locate MRPs as putative low phytate genes, and to design primers for mapping and sequencing of these putative MRP genes.

Previous studies have shown that the low phytate QTLs in CX1834 are located on LGs L and N (Walker et al., 2006; Gao et al., 2008). We mapped four soybean MIPS genes to examine their possible contribution to the low phytate phenotype based on the involvement of MIPS in another low phytate mutant, LR33 (Hitz et al., 2002). The four MIPS genes (1, 2, 3 and 4) mapped to LGs B1, G, A1, and A2, respectively. This result leads to the conclusion that the low phytate trait of CX1834 is not due to a mutation in MIPS, because none of the MIPS genes mapped to LG L or N. This corroborates the previous report that no mutations were observed in the coding regions of the four MIPS genes or non-coding regions of MIPS1 in CX1834 (Chappell et al., 2006). However, it was not known whether any mutations in non-coding regions of MIPS2, MIPS3, or MIPS4 had any association with low phytate in CX1834. The current study determines the map locations of all four MIPS genes and confirms that the low phytate trait of CX1834 is not associated with any MIPS gene, coding or non-coding, since MIPS and low phytate CX1834 QTLs do not map to the same region.

It has been shown that phytate content in maize can be controlled by several different genes (Shi et al., 2005; Shi et al., 2003; Shi et al., 2007). One class of low phytate maize mutants, *lpa1*, has a defective MRP ABC transporter, and silencing this MRP resulted in a 32-

75% reduction in phytate level in maize. Silencing of the homologous gene in soybean also resulted in reduced phytate (Shi et al., 2007). In *Arabidopsis*, the gene AtMRP5 encodes an inositol hexa*kis*phosophate transporter, and silencing it causes a low phytic acid phenotype (Nagy et al., 2009). In addition, the research reported here provides evidence that two separate single nucleotide substitutions in a similar putative MRP genes located on LGs L and N are associated with the low phytate phenotype of CX1834. The putative MRP genes on LGs L and N are the sites of the low phytate QTLs in CX1834, and LOD peaks from phytate QTL mapping correspond with the two putative MRP genes (Figure 2.2), and these MRPs provide two strong candidate genes that may be the cause of low phytate in CX1834.

Our conclusion is further supported by sequencing results from low- and high-phytate selections from a cross of CX1834 by V99-3337. All low phytate selections contained the signature CX1834 mutation of an A to a T on exon 9 of SF63MRP and a G to A mutation on exon 6 of SF165MRP. Furthermore, sequencing these exons from a set of normal phytate lines indicated the presence of the wild type sequence at the LG N mutation site, and either wild type or mutation at the LG L mutation site (Table 2.1). This is in agreement with previous studies that determined that low phytate in CX1834 is controlled by two recessive genes (Oltmans et al., 2004) corresponding to two QTLs (Walker et al. (2006); Gao et al. (2008), and that both are needed for low phytate.

The SF63MRP A to T mutation results in an amino acid change from an arginine to a stop codon (Figure 2.3), most likely causing a disruption or reduction in protein function. For the MRP on LG L, SF165MRP, the G to A mutation results in an amino acid change of an arginine to a lysine. The exact function of MRP in phytate metabolism is unknown, but may play a role in phytate storage or regulation (Raboy, 2007). These results provide evidence that

during chemical mutagenesis, M153, from which CX1834 was derived (Wilcox et al., 2000), was mutated in at least one putative MRP gene, subsequently affecting the phytate biosynthesis pathway and causing a reduction in phytate.

The results of this study provide excellent mutation-specific SNP markers for marker-assisted selection (MAS) of low phytate lines in breeding programs. As the two markers are based on the MRP genes themselves, it should facilitate MAS in crosses using CX1834 as one of the parents in large scale breeding programs.

In conclusion, we have mapped and sequenced putative MRP genes on LGs N and L that appear to contribute to low phytate phenotype in the soybean line CX1834. This A to T mutation on LG N and G to A mutation on LG L provide a SNP marker for MAS that could be useful for introgressing low phytate QTLs from CX1834 into desired breeding lines. Furthermore, this study clearly demonstrates the usefulness of the WGS in marker development and gene discovery.

ACKNOWLEDGMENTS

The study reported here was supported by the United Soybean Board.

REFERENCES

Buss G.R., Camper H.M., Roane C.W. (1988) Registration of óHutchesonô soybean. Crop Sci. 28:1024-1025.

Chappell A.S., Scaboo A.M., Wu X., Nguyen H., Pantalone V.R., Bilyeu K.D. (2006) Characterization of the MIPS gene family in *Glycine max*. Plant Breeding 125:493-500.

Churchill G.A., Doerge R.W. (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971.

Cromwell G.L., Coffey R.D., Monegue H.J., Randolph J.H. (1995) Efficacy of low-activity, microbial phytase in improving the bioavailability of phosphorus in corn-soybean meal diets for pigs. J. of Anim. Sci. 73:449-457.

Gao Y., Biyashev R.M., Saghai Maroof M.A., Glover N.M., Tucker D.M., Buss G.R. (2008) Validation of low phytate QTLs and evaluation of seedling emergence of low phytate soybeans. Crop Sci. 48:1355-1364.

Gao Y., Shang C., Saghai Maroof M.A., Biyashev R.M., Grabau E.A., Kwanyuen P., Burton J.W., Buss G.R. (2007) A modified colorimetric method for phytic acid analysis in soybean. Crop Sci. 47:1797-1803.

Gillman J.D., Pantalone V.R., Bilyeu K. (2009) The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. The Plant Genome 2:179-190.

Guttieri M., Bowen D., Dorsch J.A., Raboy V., Souza E. (2004) Identification and characterization of a low phytic acid wheat. Crop Sci. 44:418-424.

Hegeman C.E., Good L.L., Grabau E.A. (2001) Expression of D-*myo*-inositol-3-phosphate synthase in soybean. Implications for phytic acid biosynthesis. Plant Physiol. 125:1941-1948.

Hitz W.D., Carlson T.J., Kerr P.S., Sebastian S.A. (2002) Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiol. 128:650-660.

Jackson S.A., Rokhsar D., Stacey G., Shoemaker R.C., Schmutz J., Grimwood J. (2006) Toward a reference sequence of the soybean genome: A multiagency effort. Crop Sci. 46:S-55-61.

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., Newburg L. (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Larson S.R., Rutger J.N., Young K.A., Raboy V. (2000) Isolation and genetic mapping of a non-lethal rice (Oryza sativa L.) low phytic acid 1 mutation. Crop Sci. 40:1397-1405.

Larson S.R., Young K.A., Cook A., Blake T.K., Raboy V. (1998) Linkage mapping of two mutations that reduce phytic acid content of barley grain. Theor. Appl. Genet. 97:141-146.

Lott J.N.A., Greenwood J.S., Batten G.D. (1995) Mechanisms and regulation of mineral nutrient storage during seed development, in: G. G. Kigel J (Ed.), Seed Development and Germination., Marcel Dekker, New York. pp. 215ï 235.

Maughan P.J., Maroof M.A.S., Buss G.R. (2000) Identification of quantitative trait loci (QTL) affecting seed sucrose content in soybean (Glycine max). Molecular Breeding 6:105-111.

Nagy R., Grob H., Weder B., Green P., Klein M., Frelet-Barrand A., Schjoerring J.K., Brearley C., Martinoia E. (2009) The *Arabidopsis* ATP-binding cassette protein

AtMRP5/AtABCC5 is a high affinity inositol hexa*kis*phosphate transporter involved in guard cell signaling and phytate storage. Journal of Biological Chemistry 284:33614-33622.

Oltmans S.E., Fehr W.R., Welke G.A., Cianzio S.R. (2004) Inheritance of low-phytate phosphorus in soybean. Crop Sci. 44:433-435.

Raboy V. (2001) Seeds for a better future: 'low phytate' grains help to overcome malnutrition and reduce pollution. Trends Plant Sci. 6:458-462.

Raboy V. (2007) The ABCs of low-phytate crops. Nat. Biotech. 25:874-875.

Raboy V., Gerbasi P.F., Young K.A., Stoneberg S.D., Pickett S.G., Bauman A.T., Murthy P.P.N., Sheridan W.F., Ertl D.S. (2000) Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1. Plant Physiol. 124:355-368.

Rozen S., Skaletsky H. (1999) Primer3 on the WWW for General Users and for Biologist Programmers, Bioinformatics Methods and Protocols. pp. 365-386.

Saghai Maroof M.A., Biyashev R.M., Yang G.P., Zhang Q., Allard R.W. (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity chromosomal locations, and population dynamics. PNAS 91:5466-5470.

Sambrook J., Fritsch, E.F. & Maniatis, T. (1989) Molecular Cloning - A Laboratory Manual. 2nd ed. Cold Spring Habour Laboratory Press, New York.

Shi J., Wang H., Hazebroek J., Ertl D.S., Harp T. (2005) The maize low-phytic acid 3 encodes a *myo*-inositol kinase that plays a role in phytic acid biosynthesis in developing seeds. The Plant Journal 42:708-719.

Shi J., Wang H., Wu Y., Hazebroek J., Meeley R.B., Ertl D.S. (2003) The maize low-phytic acid mutant *lpa2* is caused by mutation in an inositol phosphate kinase gene. Plant Physiol. 131:507-515.

Shi J., Wang H., Schellin K., Li B., Faller M., Stoop J.M., Meeley R.B., Ertl D.S., Ranch J.P., Glassman K. (2007) Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. Nat. Biotech 25:930-937.

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S., McCouch S. (2001) Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11:1441-1452.

Van Ooijen J.W. (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

Walker D.R., Scaboo A.M., Pantalone V.R., Wilcox J.R., Boerma H.R. (2006) Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. Crop Sci. 46:390-397.

Wang S., Basten C.J., Zeng Z.-B. (2007) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm).

Wilcox J.R., Premachandra G.S., Young K.A., Raboy V. (2000) Isolation of high seed inorganic P, low-phytate soybean mutants. Crop Sci. 40:1601-1605.

Yu Y.G., Saghai Maroof M.A., Buss G.R., Maughan P.J., Tolin S.A. (1994) RFLP and microsatellite mapping of gene for soybean mosaic virus resistance. Phytopathology 84:60-64.

Yuan F.-J., Zhao H.-J., Ren X.-L., Zhu S.-L., Fu X.-J., Shu Q.-Y. (2007) Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). Theor. Appl. Genet. 115:945-957.

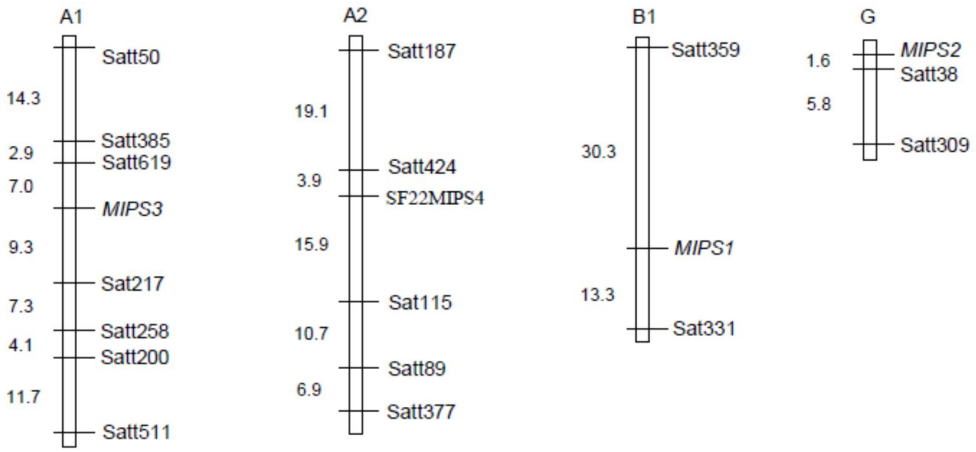


Figure 2.1. Chromosomal map positions of the MIPS gene family. Values on the left side of each linkage group are map distances in cM. Maps for LGs B1 and G were constructed based on F_2 mapping data from the CX1834-1-6 x V99-3337 population. Maps for LGs A1 and A2 were constructed based on an F_8 RIL population of V71-370 x PI407162.

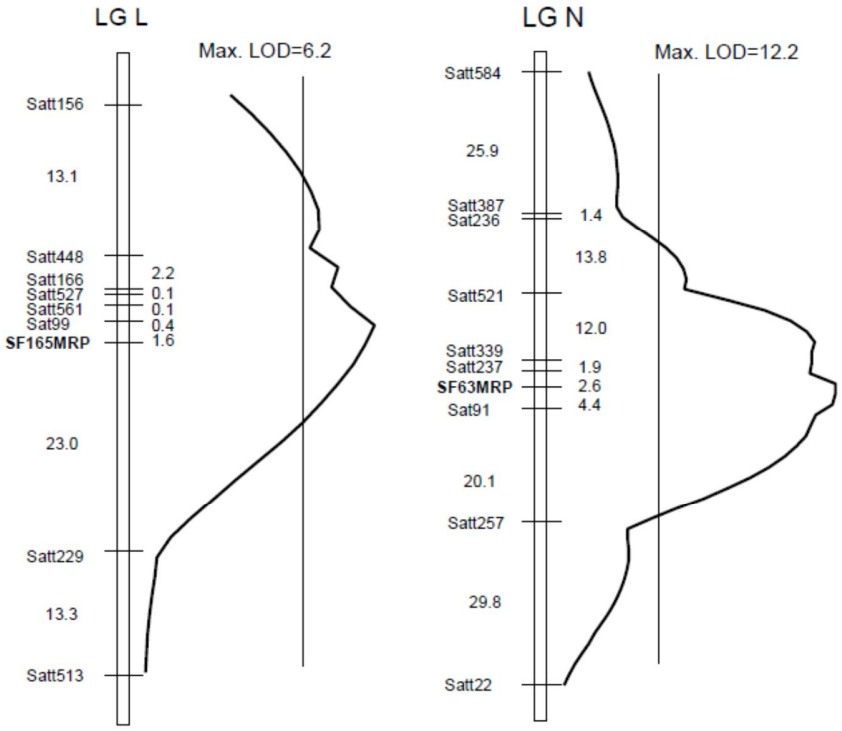


Figure 2.2. Low phytate QTL positions and LOD score plots for interval mapping on linkage groups L and N in population CX1834-1-6 x V99-3337 (modified from Gao et al. 2008). MRP gene positions are shown in bold. Map positions of MRP genes coincide with the maximum LOD score for low phytate QTLs. Solid lines parallel to the genetic map diagram indicate the LOD threshold of 4.8. Values along linkage maps are map distances between adjacent markers in cM (not to scale).

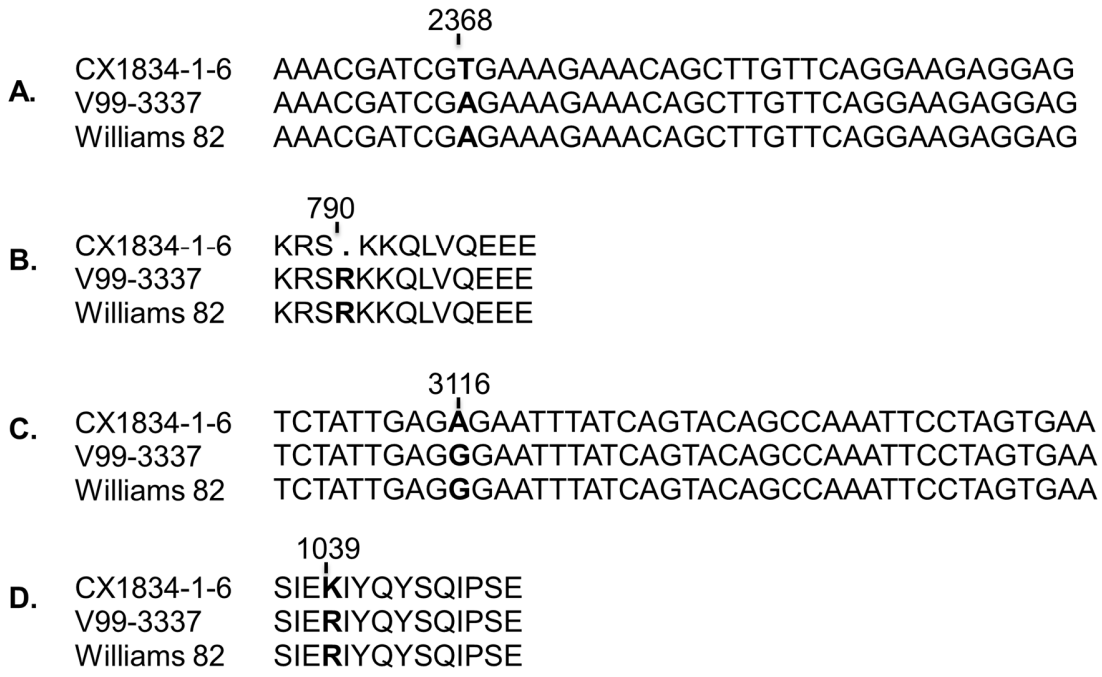


Figure 2.3. DNA and predicted amino acid sequence in soybean MRP genes on LGs N and L (SF63MRP and SF165MRP). A) Fragment of SF63MRP coding sequence alignment showing SNP at position 2368. B) Fragment of SF165MRP coding sequence alignment showing SNP at position 3116. C) Fragment of the SF63MRP predicted amino acid sequence alignment showing the result of a stop codon substitution at position 790. D) Fragment of the SF165MRP predicted amino acid sequence alignment showing the result of an arginine to lysine substitution at position 1039.

Table 2.1. Soybean lines, associated phytate content, and MRP gene sequence at position 2368 of the coding sequence. Exon 9 of SF63MRP on LG N was amplified and sequenced. For the SF63MRP gene, the mutation is a T and the wild type is an A. For the SF165MRP gene, the mutation is an A and the wild type is a G. Phytate content was determined by method of Gao et al. (2007).

Genotypes	**Phytate (mg g^{-1})**	**Nucleotide at position 2368 of SF63MRP (LG N)**	**Nucleotide at position 3115 of SF165MRP (LG L)**
CX1834-1-6ÿ§	8.6 ± 0.4	T	A
04-1272	6.2 ± 0.4	T	A
04-897	7.32	T	A
L1	5.6	T	A
L2	6.7	T	A
L3	6.8	T	A
L4	7.1	T	A
L5	7.3	T	A
L1-L5 average	6.7 ± 0.7		
H1	15.5	A	G
H2	15.6	A	A
H3	15.6	A	A
H4	16.6	A	A
H5	17.2	T	G
H1-H5 average	19.5 ± 0.6		
Athow	17.27	A	G
V99-5089ÿ§	9.9 ± 0.5	A	G
M766ÿ§	11.8 ± 0.1	A	G
Essexÿ§	13.1 ± 0.3	A	G
Hutcheson§	13.4 ± 0.2	A	G
V99-3337ÿ§	13.3 ± 0.8	A	G
V71-370ÿ§	13.3 ± 0.3	A	G
PI 87013	13.4 ± 0.1	A	G
R95-1705§	14.4 ± 0.4	A	G
Lee 68§	15.1 ± 0.1	A	G
Williams§	15.4 ± 0.4	A	G
MN1401§	15.9 ± 0.0	A	G
PI 96983§	16.0 ± 0.4	A	G
V99-8060§	16.7 ± 0.4	A	G
PI 200508ÿ§	17.9 ± 0.2	A	G
PI 407162ÿ§	19.96	A	G

ÀTable modified from Gao et al., 2008.

ÿ Lines used in polymorphism screening for SF63MRP, SF165MRP, and SF22MIPS4.

§ Lines used in polymorphism screening for MIPS genes.

Table 2.2. Primers used for MIPS gene mapping.

	GenBank Accession Number	**Primer ID**	**Primer Sequence 5ô→3ô**	**Target Motif**
MIPS1	DQ323904	MIPS-1a-F	CAAGGCTCCTCTGGTGAGTT	39 bp deletionÀ
		MIPS-1a-R	CGGTGGAACCTGTGAGAGTA	
MIPS2	DQ323906	MIPS2TACA-F	CAACGTGATTAAGGGCACAA	(TACA)7
		MIPS2TACA-R	TTTGGTGGCTTCCTTAAACG	
MIPS3	DQ323907	SPIM3-M5-F	CGATATGGTCAACAGCAACG	(T)8,2(T)7,2(T)5
		SPIM3-M5-R	TGTGTTCTTTCCACCCATGA	
MIPS4	DQ323908ÿ	SF22MIPS4-F	TTGATGCATAATAGAATAGTACAAA	(AAT)7,(TAA)4,(ATA)5,(AT)29
		SF22MIPS4-R	GAAATCATGATCAATTTTTGTTTGA	

ÀBased on deletion reported by Chappell et al. (2006).

ÿMIPS4 primer was not designed from gene sequence itself, but from region adjacent to gene.

Table 2.3. Primers designed to amplify exons of SF63MRP and SF165MRP.

LG	Exon Number	Primer ID	Primer Sequence 5ô→3ô
N	1	SF63MRP_E1-F	ATTGTTGCAAAGAAAATGCAGA
N	1	SF63MRP_E1-R	ATAGAACGTGGCACAGAAGTGA
N	2, 3	SF63MRP74/51-F	AAAGTCATACGATATCACGTATTTGAA
N	2, 3	SF63MRP74/51-R	GCATATGAGATGCCAAAATCAA
N	4	SF63MRP465F	AGTTTTGCTCTTGGGGTTTGAT
N	4	SF63MRP465R	CCTTCAACCTCTCCCAATTAGA
N	5	SF63MRP1140A-F	TCTAATTGGGAGAGGTTGAAGG
N	5	SF63MRP1140A-R	CAATAGGAACAGTGACGACGAT
N	5	SF63MRP1140B-F	TTGCTGCTATTGCAACACTGAT
N	5	SF63MRP1140B2-R	CACGCTGATTGGGAGACATA
N	6, 7	SF63MRP321F	TGGCATTTTAATCTTTTTGATTTG
N	6, 7	SF63MRP321R	GCTTGAATGTGGAAGAAAAACC
N	8, 9	SF63MRP446-F	TTTCTGTTTGTTTTCACTTTTCAA
N	8, 9	SF63MRP446-R	CTGGTGGTATACAGAGGACACA
N	10	SF63MRP165-F	GGGTAAGTTCTCTTTTGTTGTCG
N	10	SF63MRP165-R	CAGTTACAAAGATTTTAAGCTGGACA
N	11	SF63MRP295B-F	GTTTGTCTGTCACCCACCCTAT
N	11	SF63MRP295B-R	AACTGCATAAGTTGATCGTCCA
N	12	SF63MRP_E12-F	TGATGCTGAATGCAATAATGG
N	12	SF63MRP_E12-R	TGGCTTTAAAACACAAGAAAACA
N	13	SF63MRP306-F	TTTTCCCATGATCTTCTTACCC
N	13	SF63MRP306-R	CAATACCAACCTCCCAAATCTC
N	14	SF63MRP64-F	GATTAGAAGGCTGTGCATTGTG
N	14	SF63MRP64-R	GGATGGGCAAAGAAAGATACTG
N	15	SF63MRP240-F	CAGTATCTTTCTTTGCCCATCC
N	15	SF63MRP240-R	TGTTACCATACGCACCATCACT
N	16	SF63MRP_E16-F	GGCCTTGTTTCTTTCCTCTCT
N	16	SF63MRP_E16-R	CAAATTTCTCCTAGTGGGATGC
L	1	Gm19MRP_E1-F1	GTGAGAGGTTTCCAGTTTTGCT
L	1	Gm19MRP_E1-R1	ACCGCGATGGAAATTATTGTAG
L	1	Gm19MRP_E1-F2-2	AAGTCACACGAGCGGGGAGGTTGTTA
L	1	Gm19MRP_E1-R2	TGCAAAATAATCCAATGCATCTT
L	2	Gm19MRP_E2-F	TTGGTTCAGGGAAATCAAGTTT
L	2	Gm19MRP_E2-R	TATCAAATCAGCAGCAGGAAGA
L	3	Gm19MRP_E3-F	TTCCAAGGACTAGGAAGAATCC
L	3	Gm19MRP_E3-R	AGCAAACCCACCAAGTCTAAAA
L	3	Gm19MRP_E3-F-2	CCAAGGACTAGGAAGAATCC
L	3	Gm19MRP_E3-R	AGCAAACCCACCAAGTCTAAAA
L	4	Gm19MRP_E4-F	TGAATCGGGTAAGTTCTCTTCTG
L	4	Gm19MRP_E4-R	AATCAAAACATAGGGAGGGTGA
L	5	SF165MRP295-F	TCTGTCACCCTCCTATGTTTT
L	5	SF165MRP295-R	TCAGCATCATGTACAACCATCA
L	6	Gm19MRP_E6-F-2	TGATGGTTGTACATGATGCTGA
L	6	Gm19MRP_E6-R-2	ACTCCATAAAGCACCAAAGGAA
L	7	Gm19MRP_E7-F	TTGTACTGCAACTGTCTTTTACCC
L	7	Gm19MRP_E7-R	GTTTTTGGTGGGGATGTTTG
L	8,9,10	Gm19MRP_E8_E9_E10-F	AGGTTGTGCATTGTGTGATTG
L	8,9,10	Gm19MRP_E8_E9_E10-R	GGTGGCATCACTTCCTGATT

III. A POPULATION-WIDE STUDY OF SEEDLING EMERGENCE USING SOYBEAN LOW PHYTATE LINE CX1834

Abbreviations: ABC, ATP-binding cassette; ECT, extended cold test; KF, Kentland Farm; LG, linkage group; *lpa*, low phytic acid; MAS, marker-assisted selection; MH, Mt. Holly; MIPS, *myo*-inositol phosphate synthase; MRP, multidrug resistance-associated protein; P, phosphorus; PCR, polymerase chain reaction; QTL, quantitative trait loci; RIL, recombinant inbred line; SNP, single nucleotide polymorphism; SSR, simple sequence repeat

ABSTRACT

Low phytate soybeans are important from a nutritional and environmental standpoint. Unfortunately, low seedling emergence is a trait often associated with low phytate crops. The parents of the population in this study, CX1834-1-6 (hereafter CX1834) and V99-3337 have contrasting emergence and phytate values. CX1834 is a low phytate, low emerging line, and V99-3337 is a normal phytate, normal emergence line. One of the main disadvantages of the low phytate line CX1834 is its low seedling emergence. Previous studies have shown that low phytate soybean lines derived from CX1834 emerge less than normal soybean lines across multiple environments. Previous research has also shown that there were significant environmental and/or environment x phytate interactions influencing emergence. These studies were done on 10-20 normal and 10-20 low phytate lines, but to our knowledge not on a whole segregating population level. We used a population of 304 recombinant inbred lines (RILs) from a cross of CX1834 x V99-3337 for collecting genotypic and phenotypic data. The objectives of this study were to (i) determine if emergence is genotype, location (environment), year, as well as phytate dependent in this population; (ii) determine if any relationship exists between emergence and phytate level by examining their correlation coefficients, and (iii) evaluate the effectiveness of marker-assisted selection (MAS) for the low phytate trait. Phytate and emergence data were collected from seeds at two locations (Kentland Farm, Virginia and Mt. Holly, Virginia) for two years (2009 and 2010). The average phytate content of the whole CX1834 x V99-3337 RIL population was lower in 2010 than in 2009. In 2009, overall seed phytate content was lower at Kentland Farm, and in 2010, the overall phytate content was lower in Mt. Holly. In both 2009 and 2010, Mt. Holly had an overall lower emergence. Transgressive segregants for phytate and emergence were identified in this population. Phytate content was

correlated to emergence, but not strongly. Previous research has shown that low phytate in CX1834 is due to two single nucleotide mutations, one on linkage group (LG) L and the other on LG N, in two separate multidrug resistance-associated (MRP) ABC transporter genes. In all locations and years, the lines with both LG L and N MRP mutations always had the lowest emergence and phytate. This is the first study to our knowledge to evaluate emergence and phytate content on a population-wide level with a cross involving CX1834. We provide evidence that the environment significantly affects emergence. We also confirm that both LG L and N MRP mutations are needed to confer the low phytate phenotype, and that having one or the other mutations does not lower phytate any.

INTRODUCTION

Phytic acid, otherwise known as *myo*-inositol (1,2,3,4,5,6)-hexa*kis*phosphate belongs to the class of organic compounds known as inositol phosphates, having anywhere from one to six phosphate groups. Phytic acid consists of a six-membered ring with a phosphate group attached to each carbon, and is the major form of phosphorus (P) in seeds, usually representing about 75% of total seed phosphorus (Raboy, 2001). Phytic acid is known as phytate in its salt form and chelates mineral cations including calcium, iron and zinc. This phytate is stored in the protein storage vacuoles. Of the total soil P, between 29-65% is organic P (Dalai, 1977; Harrison, 1987). A differential stabilization in soils exists for organic P compounds, where the organic P compound is more strongly adsorbed to the soil as the amount of phosphate groups increases (Celi et al., 1999). Phytic acid has six phosphate groups, and therefore a high charge density, causing it to bind strongly to the soil (Anderson and Arlidge, 1962; Celi et al., 1999). Because of this strong interaction between phytic acid and the soil, phytic acid is stabilized and accumulates in soils. Therefore phytic acid makes up the majority (up to 90%) of the organic P in most soils (Harrison, 1987).

Phytic acidôs stabilization and accumulation in most soils suggests that it may be biounavailable for uptake by plants, and this may be due to its inaccessibility in the soil matrix. However, in environments where P is a limiting factor, phytic acid can be used for uptake in several species (Corona et al., 1996; Kroehler and Linkins, 1991; Li et al., 1997). There is little direct evidence showing that inositol phosphates in the soil transfer to runoff water, however, the large amount of inositol phosphates present in soil suggest there is a potential of transfer due to erosion (Turner et al., 2002). The majority of the P in animal feed is phytic acid, but unavailable to monogastric animals such as poultry, swine, and fish due to their lack of endogenous phytase,

the enzyme that breaks down phytic acid (Morse et al., 1992). Inorganic P supplements can be used to meet the animalôs dietary requirements, but the uptake is inefficient, with 70% of the total P excreted. The manure is often used as agricultural fertilizer and applied to land, increasing the potential for P transfer to runoff water (Haygarth et al., 1998). As an alternative, animal feed can be supplemented with phytase, for the breakdown of phytic acid in the animalôs gut, and this can increase P uptake efficiency and the bioavailability of certain minerals (Lei et al., 1993; Poulsen, 2000). Although phytase supplements can improve an animalôs diet, it may have adverse environmental effects. Inositol phosphates found in manure are unlikely to be transferred to runoff because they are strongly bound to the soil, but by converting phytic acid to inorganic P, phytase supplements may actually increase the risk of P transfer to runoff waters (Turner et al., 2002).

Because of its negative environmental and nutritional effects, low phytate crops are desirable (Cichy and Raboy, 2009). Several low phytate mutants have been identified and characterized in soybean (*Glycine max* L.) (Oltmans et al., 2004; Sebastian et al., 2000; Yuan et al., 2007). CX1834 is a low phytate soybean line derived from the low phytate mutant M153 (Walker et al., 2006; Wilcox et al., 2000). The low phytate trait in CX1834 is controlled by two recessive genes (Oltmans et al., 2004), and the two quantitative trait loci (QTL) controlling low phytate in CX1834 were mapped to chromosomes 19 and 3 (hereafter referred to as linkage groups (LGs) L and N, respectively) by Walker et al. (2006), and were validated in a different background by Gao et al. (2008). The low phytate phenotype in CX1834 is caused by two separate single base mutations in two multidrug resistance-associated protein (MRP) genes, homologs of the maize low phytic acid 1 *(lpa1)* gene (Gillman et al., 2009; Saghai Maroof et al.,

2009; Shi et al., 2007). These mutation-containing MRP genes are located on LGs L and N, and are the underlying cause of the low phytate QTL on these LGs.

One drawback to CX1834 is its reported low seedling emergence (Oltmans et al., 2004). Seedling emergence is a complex trait, and depends upon a variety of environmental factors including soil moisture, temperature, soil salinity, planting depth, soil impedance, and the presence of pathogens (Delouche, 1952; Hamman et al., 2002; Hegarty, 1979; Hegarty and Royle, 1978; Helms et al., 1996; Wang and Shannon, 1999). Emergence may also be influenced by genetics, as studies have reported that emergence in sweet corn is determined by multiple QTL (Han, 1994). Yousef and Juvik (2002) introgressed beneficial emergence QTLs into sweet corn commercial cultivars using marker-assisted backcrossing, and observed enhanced seedling emergence in the BC_2F_1. These introgressed QTLs were stable across genotypes and populations. Despite the reduced seedling emergence of CX1834, backcrossing CX1834 low phytate alleles on LGs L and N into a normal phytate background indicated that the reduced emergence could be overcome in some of the backcross lines and that low emergence in CX1834 may be due to other loci in the genome beside the low phytate QTLs (Spear and Fehr, 2007; Trimble and Fehr, 2009). Another study showed that the source of seeds (i.e., the environment where mother plants bearing the seeds were grown and harvested) affected the seedling emergence levels of individuals derived from CX1834 (Anderson and Fehr, 2008).

However, currently there have not been any studies evaluating the emergence on a population-wide level for crosses with CX1834 as a parent. A population-wide study of emergence would be useful for QTL discovery. The lack of this kind of study may be due to the difficulties of evaluating emergence in the field because of the strong influence of environmental effects. This inability to obtain reproducible results in field emergence tests led to the

development of the laboratory extended cold test (ECT) at Iowa State University as a way to evaluate field emergence (Spear and Fehr, 2007). Trimble and Fehr (2009) found the correlation to be as high as 92% between field emergence and emergence determined using the ECT for those low phytate lines derived from CX1834. Because of its strong correlation with field emergence, this provided us an opportunity to phenotype our population using the ECT with it still being applicable to field emergence.

Several studies have been conducted to evaluate the agronomic and seed characteristics of soybean lines with low phytate. A study by Hulke et al. (2004) used 20 low phytate and 20 normal phytate lines from a cross of CX1834 by ôB01769B019,ô a reduced palmitate, normal phytate line. Emergence and other agronomic traits were evaluated over three environments. The authors did not find any significant differences between the low phytate and normal phytate lines in yield, maturity, lodging, height, protein content, oil content, palmitate, or stearate contents. However, there was a significant difference in seedling emergence between the low phytate and normal phytate lines, with the low phytate lines emerging less (22.3% less across all environments). The authors also found a significant environment x phytate level interaction influence on emergence (Hulke et al., 2004). Another study evaluated ten low phytate and ten normal phytate lines from three crosses-- consisting of CX1834 as one parent, crossed with three different normal phytate lines (Oltmans et al., 2005). The authors found that the seedling emergence for low phytate lines was significantly lower than the normal phytate lines (23% less). There was no significant difference between the normal phytate lines and low phytate lines in terms of lodging, plant height, seed size, protein, oil, and fatty ester content. In addition, there were significant environment and environment x phytate level interactions among the three populations. Although there was a difference in seedling emergence among the low phytate and

normal phytate lines, there was enough variation to make it possible to select for lines with low phytate and improved seedling emergence. These results indicate that it is possible to develop low phytate lines with normal agronomic traits and possibly increased seedling emergence (Oltmans et al., 2005).

The objectives of the research in the current study were to observe the phytate and emergence levels of all combinations of homozygous genotypes at the mutation sites on LGs L and N in the population CX1834 x V99-3337 and to: (i) determine if emergence is genotype, location (environment), year, as well as phytate dependent in this population; (ii) determine if any relationship exists between emergence and phytate, and to (iii) evaluate the effectiveness of marker-assisted selection (MAS) for low phytate.

MATERIALS AND METHODS

Genetic Materials

The recombinant inbred population CX1834 x V99-3337 was used for marker and phenotypic data collection in this study. CX1834 is a low phytate line that was developed at Purdue University/USDA-ARS. It is derived from a cross between óAthow,ô a normal phytate soybean cultivar, and M153-1-4-6-14, a low phytate mutant. M153-1-4-6-14 descended from M153, a low phytate mutant developed by ethyl methanesulfonate mutagenesis of the soybean breeding line CX1515 (Wilcox et al., 2000). CX1834 was selected as the highest yielding low phytate line from the cross (Oltmans et al., 2004). V99-3337 is a Virginia Tech experimental line, the Roundup Ready version of óHutchesonô (Buss et al., 1988). The CX1834 x V99-3337 population has a total of 304 individuals that were used in this study. The original cross was made at Kentland Farm, Virginia, where the F_1 seeds were planted and F_1 plants selfed. Harvested F_2 seeds were planted at Kentland Farm and advanced over several generations to develop an RIL population. F_7 seed harvested in 2008 was planted at Kentland and Mt. Holly, Virginia in 2009 and harvested from both locations. F_8 seed from Kentland Farm was planted at both Kentland Farm and Mt. Holly in 2010, and the F_9 seed was harvested from both locations. Kentland Farm is approximately 150 miles west of Mt. Holly, and the soil in Kentland Farm is a Hayter loam, and the soil in Mt. Holly is a Kempsville loam (Gao et al., 2008).

In 2009, Kentland Farm had an average monthly high temperature of 27.2°C, 26.0°C, 27.9°C, 23.3°C, 18.1°C, and 14.6°C for the months of June, July, August, September, October, and November, respectively. There was a total of 48 cm of precipitation during this time (http://www.cals.vt.edu/research/kentland/collegefarm.html). Mt. Holly had an average monthly temperature high of 29.1°C, 30.1°C, 31.8°C, 26.3°C, 21.4°C, and 16.8°C for the months of June,

July, August, September, October, and November, respectively, with a total of 92 cm of precipitation (Bob Pitman, personal communication). In 2010, Kentland Farm had an average monthly temperature high of 29.6°C, 30.8°C, 29.7°C, 27.0°C, 20.8°C, and 14.5°C for the months of June, July, August, September, October, and November, respectively. There was a total of 38 cm of precipitation (http://www.cals.vt.edu/research/kentland/collegefarm.html). In 2010, Mt. Holly had an average monthly high temperature of 33.3°C, 35.1°C, 32.6°C, 31.0°C, 22.8°C, and 16.3°C for the same months, with a total of 54 cm of precipitation (Bob Pitman, personal communication).

Marker Data Collection

F_7 DNA used for marker analysis was collected from the trifoliate leaf tissue from a single plant for each line in the Kentland field in 2009. DNA extraction was performed according to (Yu et al., 1994). The locations of the MRP mutations in CX1834 on LGs L and N (Gillman et al., 2009; Saghai Maroof et al., 2009) were used to design primers for SNP genotyping. Recombinant inbred lines (n=304) from the CX1834 x V99-3337 population were genotyped for the SNP mutations on LG L and N. Two sets of primers were designed at the University of Georgia for SNP detection using melting curve analysis with the SimpleProbe assay (TIB MOLBIOL, Adelphia, NJ, USA). The SNP-L forward primer sequence is 5ô-CTGAATTTAAATGCACGTC-3ô, the reverse primer sequence is 5ô-TGTGAAGCTGAGGTTAG-3ô, and the SimpleProbe primer sequence was 5ô-TTGGCTGTACTGATA**XI**AATT**C**TCTCAATAGï Phosphate-3ô. The SNP-N forward primer sequence is 5ô-CCTGGAGGCATCTGTTATGAC-3ô, the reverse primer sequence is 5ô-CTGCCATGTATGAAAGAT-3ô, and the SimpleProbe primer sequence was 5ô-CAAGCTGTT**XI**TCTTTC**A**CGATCGTT--Phosphate-3ô. ñXIò is the internal SimpleProbe

label, and the position of the SNP in the probes is underlined and bolded. The SimpleProbes were internally labeled with fluorescein and blocked with phosphate at the 3ô end. PCR reactions were performed in 384-well plates with a total volume of 3 μL per well in a LightCycler 480 (Roche Applied Science, Indianapolis, IN, USA) following the general protocol for asymmetrical PCR. The PCR reaction mixture consisted of 20-30 ng of genomic DNA, 0.5 μM of limiting primer, 1.0 μM of excess primer, 0.2 μM of SimpleProbe, 3.0 mM $MgCl_2$, and 0.5X of LightCycler 480 Genotyping Master mix (Roche Applied Science, Indianapolis, IN, USA).

After an initial denaturation of 10 min at 95°C, 50 PCR cycles were performed with 10 s of denaturation at 95°C, 15 s of annealing at 55°C, and 20 s extension at 72°C. A final melting cycle was performed by raising the temperature to 95°C for 2 min, lowering the temperature to 40°C for 2 min, and increasing the temperature to 85°C, at this point using continuous fluorescent acquisition, and finally followed by a cool down to 40°C. The fluorescence signal (*F*) was plotted in real time against temperature (*T*) to produce melting curves for each sample. Melting curves were then converted into negative derivative curves of fluorescence with respect to temperature (-d*F*/d*T*) by the LightCycler Data Analysis software (Roche Diagnostics, Indianapolis, IN, USA). The software groups similar melting curves and automatically calls genotypes based on melting standards for parents in the experiment.

In addition to the two SNP markers from the MRP genes, data from ten SSR markers and one marker (named MIPS1a, based on a 32-bp deletion in the MIPS1 gene (Chappell et al., 2006)) were collected by capillary electrophoresis using M13 dye labeled primers on an Applied Biosystems 3130xl Genetic Analyzer (Foster City, CA) as described in Saghai Maroof et al. (2009). The markers are listed in Table 3.5; primer sequences for Satt561, Satt237, Satt339,

Satt619, AW132402, Satt424, and Satt453 are available in the BARCSOYSSR_01 soybean simple sequence repeat (SSR) database (Song et al., 2010), and primer sequences for MIPS1a, MIPS2TACA, SF63MRPD, and SF165MRPC can be found in Saghai Maroof et al. (2009). SSR marker data was visualized with the GeneMarker software version 1.60 by SoftGenetics (State College, PA). Markers were analyzed for significance with phenotypic traits using the JMP v.8 software (SAS Institute Inc., Cary, NC), with the one-way analysis of variance function. Markers were chosen based on proximity to genes known to influence phytate.

Phytate Data Collection

F_8 and F_9 seeds from the CX1834 x V99-3337 population in the field at Kentland Farm and Mt. Holly VA were harvested in 2009 and 2010. Ground powder from approximately 75 seeds from each of the 304 RILs was used in the phytate assay, conducted using a modified colorimetric method (Gao et al., 2008). Briefly, 0.5 g of soybean powder was weighed into 14 mL falcon tubes. To each sample, 10 mL of 0.65 M HCl was added, and the samples were vortexed and put on a shaker overnight (220 rpm and room temperature). Next, the samples were centrifuged at 3300 rpm at 10°C for 15 min. 500 µL of the supernatant was transferred to a microcentrifuge tube and mixed with 500 µL of 20% NaCl solution. The crude extract was allowed to precipitate for 2 hrs and then centrifuged at 13,200 rpm for 15 min. Next, 120 µL of the supernatant was added to 2.88 mL of ddH_2O (25 times dilution), and an additional 1 mL of Wadeôs Reagent (0.03% $FeCl_3 \cdot 6H_2O$ + 0.3% sulfosalicylic acid) was added for color development. The samples were then centrifuged at 3500 rpm for 10 min at 10°C. After centrifugation, the samples were read on a Beckman Coulter DU 800 Spectrophotometer (Fullerton, CA) at 500 nm. Phytate concentrations of the samples were calculated using a

calibration curve consisting of eight standards (0, 1.12, 2.24, 3.36, 5.6, 7.84, 8.96, and 11.2 ppm phytate). The eight standards also contained HCl and NaCl to eliminate matrix effects.

Emergence Data Collection

Seeds grown in 2009 and 2010 from both the Mt. Holly and the Kentland fields were sent to Iowa State Universityôs Seed Testing Laboratory (Ames, IA) for the extended cold germination test (Trimble and Fehr, 2009). Briefly, 100 seeds from each recombinant inbred line in the CX1834 x V99-3337 population were packaged and sent for the ECT assay. The seeds came from a bulk of multiple plants belonging to the same line, and were harvested from the same row in the field. The 100 seeds were planted in fiberglass trays and covered with 2.54 cm of soil. The trays were stored in a cold room at 10°C for 21 days, and then seedling emergence was calculated by counting the number of seedlings out of the 100 that had emerged after 21 days. Two subsamples were sent for emergence testing from each RIL, for each year and location. Therefore for the current study, a total of eight separate ECT experiments was performed (2009 Kentland rep 1, 2009 Kentland rep 2, 2009 Mt. Holly rep 1, 2009 Mt. Holly rep 2, 2010 Kentland rep 1, 2010 Kentland rep 2, 2010 Mt. Holly rep 1, 2010 Mt. Holly rep 2).

Statistical Analysis

Fifty-nine recombinant inbred lines that were either heterozygous or had missing SNP marker data at either of the LG L and N mutation sites were removed from the data set, since we were only interested in the homozygous genotypes. All statistical analyses including descriptive statistics and frequency distributions were performed with JMP v.8 software (SAS Institute Inc., Cary, NC). A standard least squares model was used to determine which factors were significantly influencing emergence. The two reps from each year/location were averaged to find the proportion emerged. This emergence data was tested and confirmed for normality by

plotting the residuals in a histogram and normal quantile plot. Year, location, genotypic class, phytate, year x location, location x genotypic class, year x genotypic class, and location x phytate were used as model effects. Pairwise comparison of means was based on Tukeyôs Honestly Significant Different (HSD) and Studentôs t least squared mean tests at the Ŭ = 0.05 level. Correlation coefficients for the phenotypic data were calculated with the Multivariate method in JMP. Significant p-values for quantitative trait loci were calculated using one-way ANOVA in JMP, and adjusted using Bonferroniôs method (Ŭ divided by the total number of markers tested (13)).

RESULTS AND DISCUSSION

Population-wide Emergence and Phytate Descriptive Statistics

Parental Phytate and Emergence

The means, standard deviations, and ranges were calculated for the phytate and emergence traits for both locations. These descriptive statistics for the parents of the population are listed in Table 3.1. The phytate assays for the parental lines were conducted with seed coming from each respective year and location. The CX1834 parent has significantly lower phytate than the V99-3337 parent. In 2009, CX1834 had 9.12 mg g^{-1} phytate in Kentland Farm, and 7.33 mg g^{-1} phytate in Mt. Holly. In 2009, V99-3337 had 14.23 mg g^{-1} phytate in Kentland Farm, and 12.91 mg g^{-1} phytate in Mt. Holly. In 2010, CX1834 had a phytate value of 7.22 mg g^{-1} in Kentland Farm and 8.44 in Mt. Holly, while V99-3337 had a phytate value of 13.45 mg g^{-1} in Kentland Farm, and 13.44 in Mt. Holly.

In 2009, CX1834 had much lower emergence in Mt. Holly (5%) than in Kentland Farm (43%). However, CX1834 emergence was significantly lower than V99-3337 emergence, which had the same emergence rate at both locations in 2009 (82%). In 2010, CX1834 again had much lower emergence than V99-3337; however the difference between locations for each parent was not as large. CX1834 in Kentland Farm had a mean emergence of 4%, and in Mt. Holly had a mean emergence of 2%; V99-3337 emergence in Kentland Farm was 40%, and 50% in Mt. Holly (Table 3.1).

RIL Phytate and Emergence

Individuals that were heterozygous at either LG L or N mutation site were removed. Among the RILs the phytate trait had a bimodal distribution for each year/location (Figure 3.1), with a range of phytate values from about 3-20 mg g^{-1}. The frequency distribution was skewed

towards the higher phytate values. Mean phytate values, standard deviations, and ranges for the population are shown in Table 3.1. In 2009 Kentland Farm, the mean RIL phytate for the whole population was 14.09 mg g^{-1}; in 2009 Mt. Holly the mean RIL phytate for the whole population was 15.05 mg g^{-1}. 2010 overall had lower phytate at both locations. In 2010 Kentland Farm, the mean RIL phytate for the whole population was 13.65 mg g^{-1}; in 2010 Mt. Holly, the mean RIL phytate for the whole population was 12.22 mg g^{-1}. Although these phytate means for the different environments were similar, the only means that were not significantly different from each other were 2009 Kentland and 2010 Kentland, as determined by Tukey-Kramer HSD comparison of means test at Ŭ=0.05. The differences observed in phytate content may be due to the fact that phytate is a quantitative trait under the control of numerous genes, and is most likely influenced by the environment. Although some studies have not found phytate to be influenced by the environment (Israel et al., 2006), others have shown that it is (Ahn et al., 2010; Dai et al., 2007; Maupin et al., 2011).

Overall, both Mt. Holly and Kentland Farm emergence do not appear to be normally distributed in both years (Figure 3.2). In 2009, the mean emergence of the 245 RILs was lower in Mt. Holly compared to Kentland Farm (73.9% vs. 81.7%, respectively), and the range of emergence was wider at Mt. Holly (11-97%) compared to Kentland Farm (44-96%). In 2010, Mt. Holly had lower mean emergence for the 245 RILs (27.7%) compared to Kentland Farm (30.1%); however unlike 2009, in 2010 Kentland Farm had a greater range of RIL emergence (0-87%), compared to Mt. Holly (0-70%). Mean RIL emergence from 2009 Mt. Holly, 2009 Kentland Farm, 2010 Mt. Holly, and 2010 Kentland Farm were all significantly different from one another, except 2010 Kentland Farm and 2010 Mt. Holly, which were not significantly different.

Transgressive segregants were observed for both phytate and emergence, having values either higher than V99-3337 or lower than CX1834. This indicates that there is the potential to develop higher emerging, low phytate lines compared to CX1834.

Correlations between Phytate and Emergence

To investigate the relationship between phytate and emergence, correlation coefficients were calculated for all pairwise combinations of phytate and emergence traits (Table 3.2), and all correlations were found to be positive at the $p<0.0001$ significance level. (Except between 2009 Kentland emergence and 2010 Kentland phytate data, which had $p=0.0003$ significance level.) In 2009, Kentland Farm phytate was correlated with Kentland Farm emergence, although it was not strong, at 0.35. 2009 Mt. Holly phytate had a higher correlation with 2009 Mt. Holly emergence, at 0.45. In 2010, Kentland Farm phytate was correlated with Kentland Farm emergence at 0.42. 2010 Mt. Holly phytate was correlated with Mt. Holly emergence with a correlation coefficient of 0.31. These low correlations between phytate and emergence suggest that low phytate does not necessarily indicate low emergence. All phytate data sets (2009 Mt. Holly, 2009 Kentland Farm, 2010 Mt. Holly, 2010 Kentland Farm) were highly correlated with each other with correlation coefficients ranging from 0.76 to 0.86 (Table 3.2). All emergence data sets were significantly correlated with each other at the $p<0.0001$ level as well, however, these correlations were not strong and the values ranged from 0.27 to 0.49 (Table 3.2). Low correlations in emergence between the same year and different locations, as well as between the same location and different years suggest that emergence is not stable. For a given line, low emergence for one year or in one location does not necessarily predict the next year or a different location.

Phytate, Genotype, Environment, and its Effect on Emergence

Because of their known association with low phytate in CX1834 (Gillman et al., 2009; Saghai Maroof et al., 2009), we used the LG L and N SNP mutations to design perfect markers for the MRP genes and to genotype the whole CX1834 x V99-3337 RIL population. These perfect markers could be extremely valuable for MAS in breeding programs, as they are based on the causative mutations in CX1834 which confer low phytate. All individuals were grouped into one of four classes, representing the various combinations of the two mutant alleles. The four classes are as follows, designated ñgenotype at L mutation site/genotype at N mutation siteò: CX/CX, CX/V, V/CX, and V/V. ñCXò represents an allele homozygous for the mutation, and ñVò represents the homozygous wild type allele from V99-3337. When comparing the different genotypic classôs phytate levels, individuals with the CX/CX genotype (homozygous for both L and N mutations) had the lowest phytate levels averaged over all years and locations, with a mean of 7.21 mg g^{-1}, which was significantly lower than all other classes. The CX/V class mean was 15.48 mg g^{-1}, V/CX was 15.01 mg g^{-1}, and V/V was 15.51 mg g^{-1} (Figure 3.3). The phytate values of these three classes, each having at least one wild type allele for the LG L or N mutation, were not significantly different from each other.

These results are in agreement with previous studies (Gillman et al., 2009; Oltmans et al., 2004; Walker et al., 2006) that have indicated that the plant must be homozygous for both mutations to have phytate levels matching or lower than CX1834. This is also in agreement with the study by Scaboo et al. (2009), which shows that in a population, those individuals belonging to the genotypic class with the homozygous CX1834 allele at both Satt237 and Satt561 (markers linked to the L and N MRP genes) loci had the lowest phytate levels. In the current study, the three categories containing at least one wild-type allele were virtually identical for phytate

content, and this indicates that neither L mutation nor N mutation has a greater effect on lowering phytate.

In order to determine which main effects were significantly affecting emergence we fit a standard least squares model to combined data for the four environments. The resulting model which predicts emergence had a summary of fit R^2 value of 0.69. The results are shown in Table 3.3, and reveal that variation among year, location, genotypic class, year x genotypic class, and year x location interactions were all significant for emergence with an F test p-value Ò 0.0002. Phytate and location x genotypic class were not shown to significantly affect emergence.

The standard least squares model however indicated that genotypic class was significantly associated with emergence level ($p < 0.0001$). When looking at the emergence data over both years and locations and broken down by genotypic class, CX/CX did have the lowest mean emergence (42%), and this was significantly different from the other genotypic classes (Figure 3.4). CX/V had a mean emergence of 56%, and V/CX and V/V had a mean emergence of 57% over both years and locations. Upon examining the emergence data broken down by year and genotypic class, it is clear that over all genotypes and all locations, the emergence in 2010 was lower than in 2009 (Table 3.4). Over all years and locations the CX/CX genotypic class had the lowest emergence when considering the individual sets of emergence data. In every environment, the CX/CX mean emergence was significantly lower than the other genotypic classes (CX/V, V/CX, and V/V) (Table 3.4). The low emergence of the CX/CX class in 2010 (10% in Kentland Farm and 14% in Mt. Holly), as well as the relatively low emergence of all the other genotypic classes in 2010 compared to 2009, could possibly be explained by the different environmental conditions between 2009 and 2010. 2009 had a daily high temperature average of 24.4°C, and a total precipitation of 140 cm (sum of both locations) for the months of

June through November. 2010 had a daily high temperature average of 27.0°C, with a sum of 92 cm of precipitation for the same months. The difference in temperature and less rain in 2010 could have been what caused the overall lower emergence in 2010 compared to 2009.

Molecular Markers Associated with Seed Phytate and Emergence

There are four *myo*-inositol phosphate synthase (MIPS) genes in soybean (Chappell et al., 2006; Hegeman et al., 2001), located on LGs A1, A2, B1, and G (Saghai Maroof et al., 2009). MIPS genes are known to be the sole source of *myo*-inositol in the phytic acid pathway, and to be associated with low phytate in some plants (Hitz et al., 2002; Sebastian et al., 2000). However, it is unknown if MIPS plays a role in seedling emergence. In order to test this, we either designed primers from the MIPS genes themselves (Saghai Maroof et al., 2009) or used publicly available SSR markers located near the MIPS genes in order to determine if the MIPS genes were associated with emergence (Table 3.5). The SNP markers from the LG L and N MRPs were also tested for association.

We used the 2009 Kentland Farm phytate data to test for marker trait association. All sets of emergence data (both years and both locations) were tested for significance with MIPS markers using one-way ANOVA. None were found to be significant except Satt424 (3.9 cM from MIPS2) with 2009 Kentland Farm emergence and 2010 Mt. Holly emergence, and the marker MIPS2TACA (from the MIPS2 gene), at the $p < 0.05$ significance level. This may indicate the possibility of MIPS2 or a gene located near MIPS2 to be influencing emergence. However, these QTLs did not appear to be stable over years or locations, and may be heavily influenced by environment.

In order to determine whether any other regions of known low phytate QTLs or known phytate-related genes are associated with emergence, we tested several other markers in specific

regions of interest. The first markers initially tested are located in the low phytate QTL regions of LG L or N and are known to be tightly linked to MRP genes. These include the SSR markers SF165MRPC and Satt561 for LG L, and SF63MRPD, Satt237, and Satt339 from LG N (Saghai Maroof et al., 2009; Scaboo et al., 2009; Walker et al., 2006). In addition, the SNP markers from the LG L and N MRPs were tested.

In contrast to the MIPS markers, nearly all of the markers associated with the LG L and N MRP mutations were significant with all emergence data sets, except the SNP-N and Satt339 marker with 2010 Mt. Holly emergence. However, the majority of the markers were significant with emergence at the $p < 0.001$ probability level. These markers were also significantly associated with phytate, which is to be expected. A genome-wide study to indentify emergence QTLs is currently underway in order to identify unknown regions of the genome that may be influencing emergence.

The results (Table 3.5) show a strong association between the LG L and N SSR and SNP markers near the MRP genes and phytate, which has been previously reported (Gao et al., 2008; Scaboo et al., 2009; Walker et al., 2006). Also, a strong association of these markers with 2009 emergence in both Kentland Farm and Mt. Holly was detected (Table 3.5). Because the main low phytate QTLs in CX1834 are located on LGs L and N due to mutations in an MRP gene, it is logical that the same regions would show significance with emergence.

Table 3.6 displays the selection efficiency of several markers for correctly identifying low phytate individuals in the population. The L and N SNP markers had a very high efficiency, correctly identifying 97.9% of the individuals that carried both low phytate mutations as being low phytate. This is better than the microsatellite markers, which only had a 90% correctly selected individuals when using SF165MRP/SF63MRP, or 88.7% when using Satt237/Satt561.

The lowest phytate lines (<7.5 mg g^{-1}) all had the CX/CX genotype. MAS using the L and N SNP markers would be more successful in identifying low phytate individuals than using SSR markers, and these markers could be of use in breeding programs for selecting low phytate lines. This also gives further evidence that the L and N mutations in the MRP genes are indeed the causative mutations conferring low phytate in CX1834, in agreement with previous studies (Gillman et al., 2009; Saghai Maroof et al., 2009).

CONCLUSIONS

The parents of the population in this study, CX1834 and V99-3337 have contrasting emergence and phytate values. CX1834 is a low phytate, low emerging line, and V99-3337 is a normal phytate, normal emergence line. For phytate of the whole RIL population, the overall RIL phytate was lower in 2010 than in 2009. In 2009, overall phytate was lower in Kentland Farm, and in 2010, the overall phytate was lower in Mt. Holly. In both 2009 and 2010, Mt. Holly had an overall lower emergence. Transgressive segregants for the phytate and emergence traits were identified in this population. Phytate was correlated to emergence, but not strongly. In all locations and years, the CX/CX genotypic class always had the lowest emergence and phytate. This confirms that you need both mutations to be low phytate. We tested a few SSR markers for association with phytate and emergence but did not find any new significant associations. We are currently planning on doing a genome-wide SNP analysis in order to search for other phytate and/or emergence QTL.

REFERENCES

Ahn D.J., Won J.G., Rico C.M., Lee S.C. (2010) Influence of variety, location, growing year, and storage on the total phosphorus, phytate-phosphorus, and phytate-phosphorus to total phosphorus ratio in rice. Journal of Agricultural and Food Chemistry 58:3008-3011.

Anderson B.P., Fehr W.R. (2008) Seed source affects field emergence of low-phytate soybean lines. Crop Sci. 48:929-932.

Anderson G., Arlidge E.Z. (1962) The adsorption of inositol phosphates and glycerophosphate by soil clays, clay minerals, and hydrated sesquioxides in acid media. Journal of Soil Science 13:216-224.

Buss G.R., Camper H.M., Roane C.W. (1988) Registration of ôHutchesonô soybean. Crop Sci. 28:1024-1025.

Celi L., Lamacchia S., Marsan F.A., Barberis E. (1999) Interaction of inositol hexaphosphate on clays: adsorption and charging phenomena. Soil Science 164:574-585.

Chappell A.S., Scaboo A.M., Wu X., Nguyen H., Pantalone V.R., Bilyeu K.D. (2006) Characterization of the MIPS gene family in *Glycine max*. Plant Breeding 125:493-500.

Cichy K.A., Raboy V. (2009) Evaluation and development of low phytate crops, in: H. Krishnan (Ed.), Modification of Seed Composition to Promote Health and Nutrition, American Society of Agronomy and Crop Science Society of America, Madison, WI. pp. 177-201.

Corona M.E.P., Klundert I.v.d., Verhoeven J.T.A. (1996) Availability of organic and inorganic phosphorus compounds as phosphorus sources for *Carex* species. New Phytologist 133:225-231.

Dai F., Wang J., Zhang S., Xu Z., Zhang G. (2007) Genotypic and environmental variation in phytic acid content and its relation to protein content and malt quality in barley. Food Chemistry 105:606-611.

Dalai R.C. (1977) Soil Organic Phosphorus, in: N. C. Brady (Ed.), Advances in Agronomy, Academic Press. pp. 83-117.

Delouche J.C. (1952) Influence of moisture and temperature levels on the germination of corn, soybeans, and watermelon. Proc. Assoc. Off. Seed Anal. 43:117-126.

Gao Y., Biyashev R.M., Saghai Maroof M.A., Glover N.M., Tucker D.M., Buss G.R. (2008) Validation of low phytate QTLs and evaluation of seedling emergence of low phytate soybeans. Crop Sci. 48:1355-1364.

Gillman J.D., Pantalone V.R., Bilyeu K. (2009) The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. The Plant Genome 2:179-190.

Hamman B., Egli D.B., Koning G. (2002) Seed vigor, soilborne pathogens, preemergent growth, and soybean seedling emergence. Crop Sci. 42:451-457.

Han T. (1994) Investigations into the physiology and genetics of sweet corn seedling emergence and vigor. MS thesis, Univ. of Illinois, Urgana-Champaign.

Harrison A.F. (1987) Soil organic phosphorus. A review of world literature. CAB International., Oxford.

Haygarth P.M., Chapman P.J., Jarvis S.C., Smith R.V. (1998) Phosphorus budgets for two contrasting grassland farming systems in the UK. Soil Use and Management 14:160-167.

Hegarty T.W. (1979) Factors influencing the emergence of calabrese and carrot seedlings in the field. J. Hortic. Sci. 54:199-207.

Hegarty T.W., Royle S.M. (1978) Soil impedance as a factor reducing crop seedling emergence, and its relation to soil conditions at sowing, and to applied water. Journal of Applied Ecology 15:897-904.

Hegeman C.E., Good L.L., Grabau E.A. (2001) Expression of D-*myo*-inositol-3-phosphate synthase in soybean. Implications for phytic acid biosynthesis. Plant Physiol. 125:1941-1948.

Helms T.C., Deckard E.L., Goos R.J., Enz J.W. (1996) Soil moisture, temperature, and drying influence on soybean emergence. Agron J 88:662-667.

Hitz W.D., Carlson T.J., Kerr P.S., Sebastian S.A. (2002) Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiol. 128:650-660.

Hulke B.S., Fehr W.R., Welke G.A. (2004) Agronomic and seed characteristics of soybean with reduced phytate and palmitate. Crop Sci. 44:2027-2031.

Israel D.W., Kwanyuen P., Burton J.W. (2006) Genetic variability for phytic acid phosphorus and inorgaic phosphorus in seeds of soybeans in maturity groups V, VI, and VII. Crop Sci. 46:67-71.

Kroehler C.J., Linkins A.E. (1991) The absorption of inorganic phosphate from ^{32}P-labeled inositol hexaphosphate by *Eriophorum vaginatum*. Oecologia 85:424-428.

Lei X., Ku P.K., Miller E.R., Ullrey D.E., Yokoyama M.T. (1993) Supplemental microbial phytase improves bioavailability of dietary zinc to weanling pigs. The Journal of Nutrition 123:1117-1123.

Li M., Osaki M., Madhusudana Rao I., Tadano T. (1997) Secretion of phytase from the roots of several plant species under phosphorus-deficient conditions. Plant and Soil 195:161-169.

Maupin L.M., Rosso M.L., Rainey K.M. (2011) Environmental effects on soybean with modified phosphorus and sugar composition. Crop Sci. 51:642-650.

Morse D., Head H.H., Wilcox C.J. (1992) Disappearance of phosphorus in phytate from concentrates in vitro and from rations fed to lactating dairy cows. Journal of Dairy Science 75:1979-1986.

Oltmans S.E., Fehr W.R., Welke G.A., Cianzio S.R. (2004) Inheritance of low-phytate phosphorus in soybean. Crop Sci. 44:433-435.

Oltmans S.E., Fehr W.R., Welke G.A., Raboy V., Peterson K.L. (2005) Agronomic and Seed Traits of Soybean Lines with Low-Phytate Phosphorus. Crop Sci. 45:593-598.

Poulsen H.D. (2000) Phosphorus utilization and excretion in pig production. Journal of Environmental Quality 29:24-27.

Raboy V. (2001) Seeds for a better future: 'low phytate' grains help to overcome malnutrition and reduce pollution. Trends Plant Sci. 6:458-462.

Saghai Maroof M.A., Glover N.M., Biyashev R.M., Buss G.R., Grabau E.A. (2009) Genetic basis of the low-phytate trait in the soybean line CX1834. Crop Sci. 49:69-76.

Scaboo A.M., Pantalone V.R., Walker D.R., Boerma H.R., West D.R., Walker F.R., Sams C.E. (2009) Confirmation of molecular markers and agronomic traits associated with seed phytate content in two soybean RIL populations. Crop Sci. 49:426-432.

Sebastian S.A., Kerr P.S., Pearlstein R.W., Hitz W.D. (2000) Soybean germplasm with novel genes for improved digestibility, in: J. K. Drackley (Ed.), Soy in Animal Nutrition, Federation of Animal Science Societies, Savoy, IL. pp. 56-74.

Shi J., Wang H., Schellin K., Li B., Faller M., Stoop J.M., Meeley R.B., Ertl D.S., Ranch J.P., Glassman K. (2007) Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. Nat. Biotech 25:930-937.

Song Q., Jia G., Zhu Y., Grant D., Nelson R.T., Hwang E.-Y., Hyten D.L., Cregan P.B. (2010) Abundance of SSR motifs and development of candidate polymorphic SSR Markers (BARCSOYSSR_1.0) in soybean. Crop Sci. 50:1950ï 1960.

Spear J.D., Fehr W.R. (2007) Genetic improvement of seedling emergence of soybean lines with low phytate. Crop Sci. 47:1354-1360.

Trimble L.A., Fehr W.R. (2009) Genetic improvement of seedling emergence of low-phytate soybean lines. Crop Sci. 50:67-72.

Turner B.L., Papházy M.J., Haygarth P.M., Mckelvie I.D. (2002) Inositol phosphates in the environment. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357:449-469.

Walker D.R., Scaboo A.M., Pantalone V.R., Wilcox J.R., Boerma H.R. (2006) Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. Crop Sci. 46:390-397.

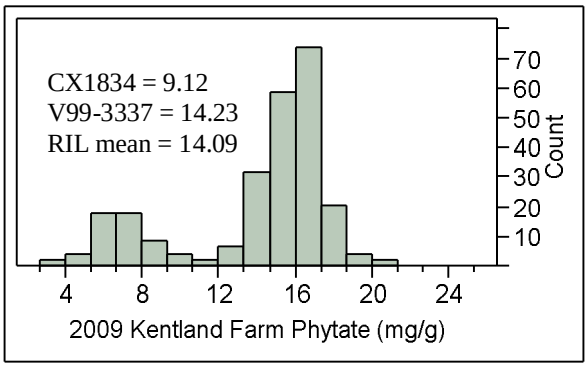
Wang D., Shannon M.C. (1999) Emergence and seedling growth of soybean cultivars and maturity groups under salinity. Plant and Soil 214:117-124.

Wilcox J.R., Premachandra G.S., Young K.A., Raboy V. (2000) Isolation of high seed inorganic P, low-phytate soybean mutants. Crop Sci. 40:1601-1605.

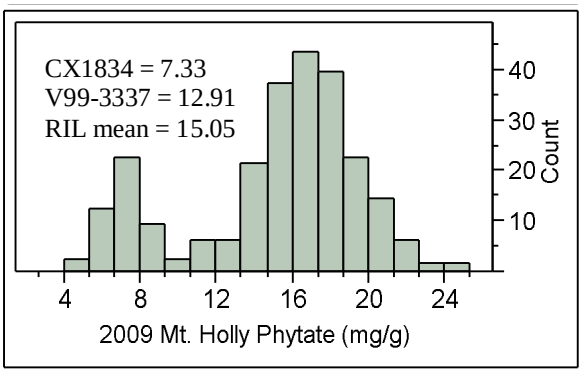
Yousef G.G., Juvik J.A. (2002) Enhancement of seedling emergence in sweet corn by marker-assisted backcrossing of beneficial QTL. Crop Sci. 42:96-104.

Yu Y.G., Saghai Maroof M.A., Buss G.R., Maughan P.J., Tolin S.A. (1994) RFLP and microsatellite mapping of gene for soybean mosaic virus resistance. Phytopathology 84:60-64.

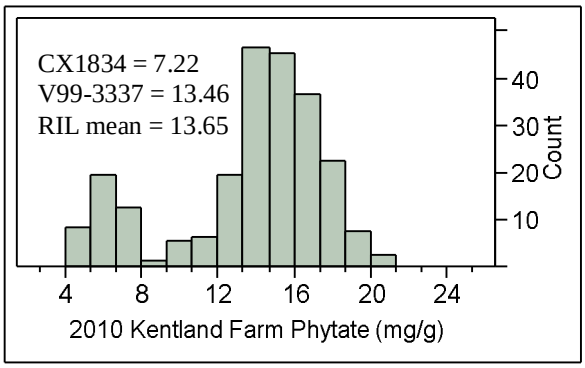
Yuan F.-J., Zhao H.-J., Ren X.-L., Zhu S.-L., Fu X.-J., Shu Q.-Y. (2007) Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). Theor. Appl. Genet. 115:945-957.



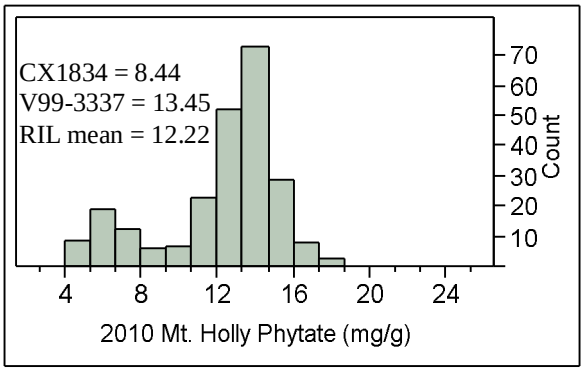
A



B

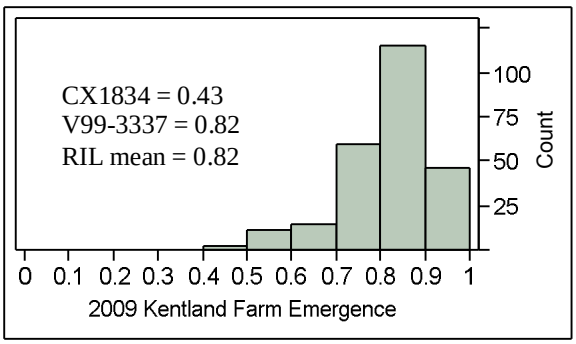


C

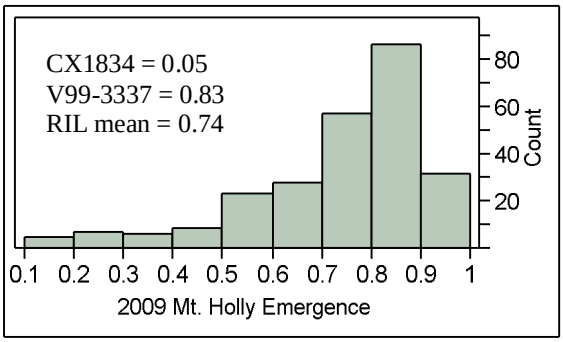


D

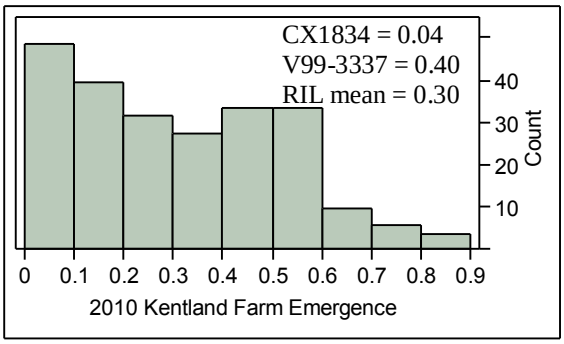
Figure 3.1. Frequency distribution for seed phytate (mg g^{-1}) of 245 RILs from the F_8 and F_9 generations of the CX1834 x V99-3337 population from (A) Kentland Farm in 2009, (B) Mt. Holly in 2009, (C) Kentland Farm in 2010, and (D) Mt. Holly in 2010. Phytate values for each parent at each environment are displayed on the graph.



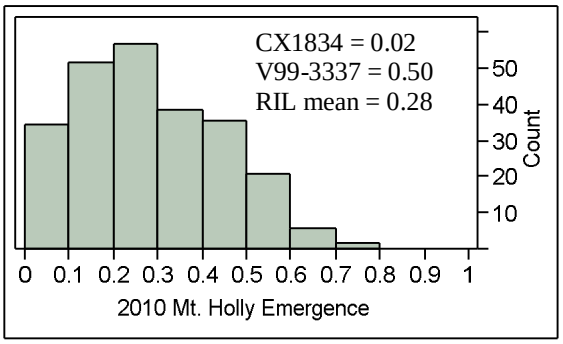
A



B



C



D

Figure 3.2. Frequency distribution for mean emergence (proportion emerged) of 245 RILs from the cross CX1834 x V99-3337 from (A) Kentland Farm in 2009, (B) Mt. Holly in 2009, (C) Kentland Farm in 2010, and (D) Mt. Holly in 2010. Emergence values for each parent at each environment are displayed on the graph.

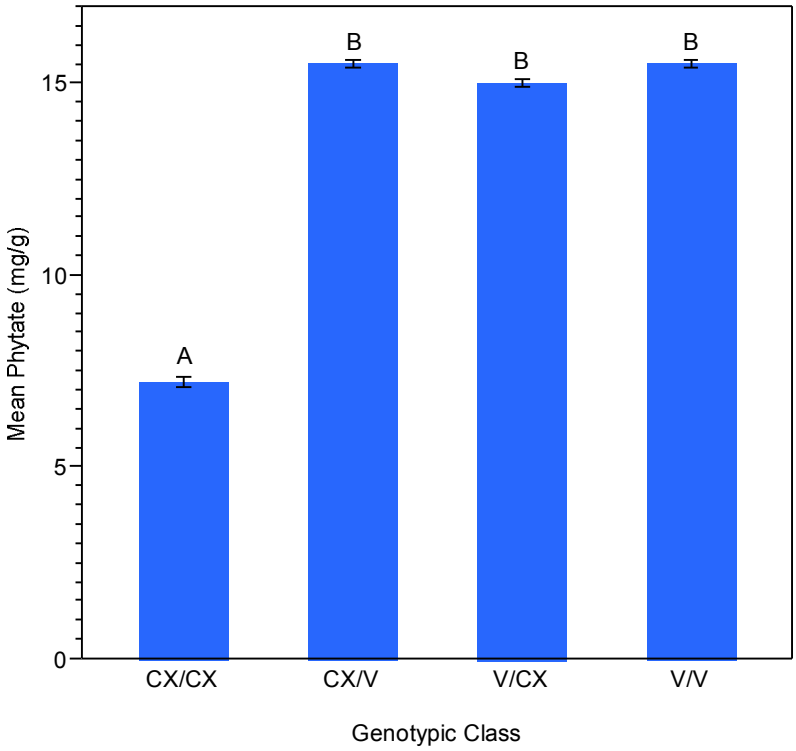


Figure 3.3. Mean phytate concentration for all four homozygous genotypic classes for the population CX1834 x V99-3337, consisting of 245 RILs. Phytate values are averaged from 2009 and 2010 Kentland Farm and Mt. Holly seed, and genotypic classes are based on the LG L and N SNP markers at the causative mutation sites. Genotypic classes are represented as ñallele at L mutation site/allele at N mutation site,ò with ñCXò representing the mutant CX1834 allele and ñVò representing the V99-3337 wild-type allele. Over all years and locations, CX1834 had a mean phytate level of 8.03 mg $g^{-1}$and V99-3337 had a mean phytate level of 13.51 mg g^{-1}. Each error bar is constructed using 1 standard error from the mean. Levels not connected by the same letter are significantly different based on Tukey-Kramer HSD comparison of means at an Ŭ of 0.05.

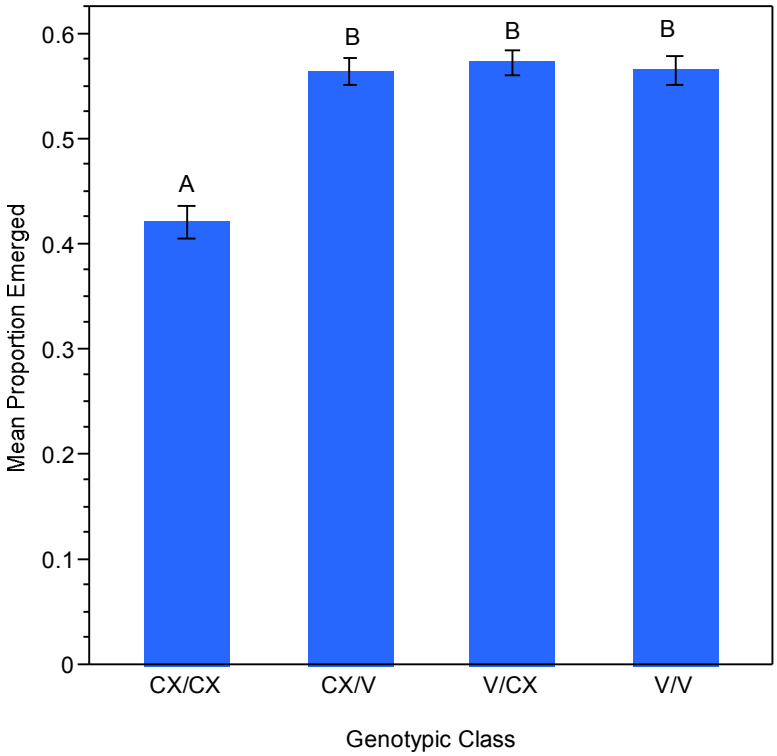


Figure 3.4. Mean emergence for all four homozygous genotypic classes for the population CX1834 x V99-3337, consisting of 245 RILs. Emergence values are averaged from the combined data sets of 2009 Kentland, 2009 Mt. Holly, 2010 Kentland, and 2010 Mt. Holly. Genotypic classes are based on the LG L and N SNP markers at the causative mutation site. Genotypic classes are represented as ñallele at L mutation site/allele at N mutation site,ò with ñCXò representing the mutant CX1834 allele and ñVò representing the V99-3337 wild-type allele. Over all years and locations, CX1834 had a mean emergence of 14% and V99-3337 had a mean emergence of 64%. Each error bar is constructed using 1 standard error from the mean. Levels not connected by the same letter are significantly different based on Tukey-Kramer HSD comparison of means at an Ŭof 0.05.

Table 3.1. Descriptive statistics for emergence and phytate traits in 245 RILs from the cross CX1834 x V99-3337 from Kentland Farm, VA (KF), and Mt. Holly, Virginia (MH) in 2009 and 2010.

	Phytate (mg g^{-1})						**Emergence (%)**					
	2009			**2010**			**2009**			**2010**		
PARENTS	Mean		SD[§]	Mean		SD	Mean		SD	Mean		SD
CX1834												
KF	9.12		1.24	7.22		2.86	42.9		2.6	4.0		3.0
MH	7.33		1.08	8.44		0.83	5.3		6.4	2.0		2.0
V99-3337												
KF	14.23		0.85	13.46		0.25	82.4		17.8	40.5		10.5
MH	12.91		0.35	13.45		1.22	82.7		7.9	50.0		16.0
RILs	Mean[A]	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
KF	14.09B	3.89	3.86-20.14	13.65B	3.94	4.09- 21.03	81.7A	9.3	44-96	30.1C	21.1	0-87
MH	15.05A	4.41	4.74-24.0	12.22C	3.10	4.42- 18.62	73.9B	16.9	11-97	27.7C	16.3	0-70

[A]For means, levels not connected by the same letter are significantly different, based on Tukey-Kramer HSD comparison of means at an Ŭof 0.05.

[§] SD = standard deviation

Table 3.2. Phenotypic correlation coefficients for phytate and emergence traits for both locations and years for the 245 RILs from the CX1834 x V99-3337 population. All correlations are significant at the p=0.001 level.

Trait	2009 MH Phytate	2010 KF Phytate	2010 MH Phytate	2009 KF Emerg.	2009 MH Emerg.	2010 KF Emerg.	2010 MH Emerg.
2009 KF Phytate	0.86	0.84	0.83	0.35	0.55	0.41	0.36
2009 MH Phytate		0.79	0.75	0.28	0.45	0.41	0.39
2010 KF Phytate			0.78	0.25	0.47	0.42	0.37
2010 MH Phytate				0.31	0.45	0.36	0.31
2009 KF Emergence					0.46	0.27	0.27
2009 MH Emergence						0.43	0.49
2010 KF Emergence							0.45

Table 3.3. A. Analysis of Variance for the Standard Least Squares model using proportion emerged as the response. B. Model effect tests of sources of variation on the proportion emerged of 245 RILs grown in 2009 and 2010 at Mt. Holly and Kentland Farm.

A

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Model	14	115.12	8.22	296.52
Error	1828	50.69	0.03	**Prob > F**
C. Total	1842	165.82		0.0000

B

Source	Degrees of Freedom	Sum of Squares	F Ratio	Prob > F
year	1	88.4040	3187.835	0.0000
location	1	1.2264	44.2231	<0.0001
genotypic class	3	1.5907	19.1204	<0.0001
phytate	1	0.0429	1.5499	0.2133
year*location	1	0.4389	15.8281	<0.0001
location*genotypic class	3	0.1216	1.4622	0.2230
year*genotypic class	3	0.4547	5.4652	0.0010
location*phytate	1	0.1431	5.1604	0.0232

Table 3.4. Mean emergence for all four genotypic classes, broken down by year/location. Levels not connected by the same letter are significantly different, based on Tukey-Kramer HSD comparison of means at an Ŭof 0.05.

	Mean Emergence (%)			
Genotypic class	2009 KF	2009 MH	2010 KF	2010 MH
CX/CX	74 A	52 A	10 A	14 A
CX/V	83 B	78 B	34 B	28 B
V/CX	83 B	79 B	38 B	33 B
V/V	83 B	80 B	32 B	31 B

Table 3.5. Quantitative trait loci determined by one-way analysis of variance for phytate and emergence using 245 RILs of the population CX1834 x V99-3337. 2009 Kentland Farm (KF) phytate data was used, and mean emergence (averaged from both reps) for each year and location. P-values for each marker-trait combination are listed, and significant p-value cutoffs were calculated by Bonferroniôs method (Ždivided by the number of markers tested).

Marker[A]	**Distance away from gene of interest (bp)**	**LG**	**Phytate**	**2009 KF Emerg. (mean)**	**2009 MH Emerg. (mean)**	**2010 KF Emerg. (mean)**	**2010 MH Emerg. (mean)**
SNP-L	from MRP gene	L	0.000***	0.001*	0.000***	0.000***	0.000**
SF165MRPC	526 bp	L	0.000***	0.000*	0.000***	0.000***	0.000**
Satt561	254,122 bp	L	0.000***	0.000*	0.000***	0.000***	0.000**
SNP-N	from MRP gene	N	0.000***	0.001*	0.000***	0.030	0.089
SF63MRPD	from MRP gene	N	0.000***	0.000**	0.000***	0.006*	0.045
Satt237	97,440 bp	N	0.000***	0.000**	0.000***	0.029	0.024
Satt339	281,980 bp	N	0.000***	0.000**	0.000**	0.034	0.127
Satt619	878,438 bp	A1	0.469	0.479	0.664	0.275	0.713
AW132402	1,146,795 bp	A2	0.653	0.878	0.763	0.984	0.229
Satt424	36,602 bp	A2	0.09	0.016	0.75	0.612	0.023
MIPS2TAC	from MIPS2 gene	G	0.364	0.555	0.324	0.790	0.033
MIPS1a	from MIPS1 gene	B1	0.687	0.533	0.142	0.087	0.381
Satt453	unknown	B1	0.238	0.802	0.537	0.332	0.969

ÀSF165MRPC and SF63MRPD are SSR markers from within or near the LG L and LG N MRP genes, respectively. MIPS2TACA is an SSR marker designed from the MIPS2 gene, and MIPS1a is an indel marker based on a deletion in the MIPS1 gene. Primer sequences can be found in (Saghai Maroof et al., 2009).

*Significance at the Ž=0.05 probability level

**Significance at the Ž=0.01 probability level

***Significance at the Ž=0.001 probability level

Table 3.6. Selection efficiency of SNP markers and microsatellites with either one or both low phytate alleles coming from L and N MRP markers (and others tightly linked). Selection efficiency was designated as the % of lines having the low phytate alleles for a particular marker that actually displayed the low phytate phenotype.

	SNP-L	SNP-N	Both	Low phytate lines
	119	121	48	47
% correctly selected	39.5%	38.8%	97.9%	
	SF165MRP	SF63MRP	Both	Low phytate lines
	121	139	51	46
% correctly selected	38.0%	33.1%	90.2%	
	Satt561	Satt237	Both	Low phytate lines
	126	141	53	47
% correctly selected	37.3%	33.3%	88.7%	

IV. A STUDY OF TWO LOW PHYTATE GENES AND THEIR INTERACTION IN THE RECOMBINANT INBRED POPULATION V99-5089 x CX1834

Abbreviations: ABC, ATP-binding cassette; HPLC, high performance liquid chromatography; LG, linkage group; MIPS, *myo*-inositol phosphate synthase; MRP, multidrug resistance-associated protein; P, phosphorus; Pi, inorganic phosphorus; PSV, protein storage vacuole; RFO, raffinose family oligosaccharide; RIL, recombinant inbred line; SNP, single nucleotide polymorphism

ABSTRACT

Low phytate, low raffinosaccharide soybeans [*Glycine max* (L.) Merr.] are desirable from a nutritional and environmental standpoint. The low phytate mutant CX1834-1-6 is known to have two SNP-type mutations in two separate multidrug resistance-associated protein (MRP) ABC transporters located on linkage groups (LGs) L and N which are responsible for the low phytate phenotype. Another low phytate line, V99-5089, also has low levels of raffinosaccharides, making it very valuable in breeding programs. Low phytate and raffinosaccharide content in V99-5089 has been shown to be caused by another SNP-type mutation in its *myo*-inositol phosphate synthase (MIPS1) gene on LG B1. The recombinant inbred population CX1834-1-6 x V99-5089 (n=173) provides the opportunity to study the effect of different combinations of all three mutations on phytate and oligosaccharide levels. The lines with the lowest average phytate content (6.72 mg g^{-1}; n=14) were those with all three mutations, but were not significantly lower than the lines with only the two MRP mutations originating from CX1834-1-6 (6.94 mg g^{-1}; n=17). However, looking at the oligosaccharide content presents a different story. Only the B1 MIPS mutation confers low stachyose - except in the case of all three mutations combined. Although this genotypic class has the MIPS mutation and was expected to be low stachyose, combining it with the MRP mutations caused low phytate and high stachyose instead of low phytate and low stachyose. It is not entirely clear why this occurs, but the phytate and raffinosaccharide biochemical pathways are related with many common intermediates. One explanation could be that because phytate cannot be transported into the protein storage vacuoles (due to the MRP mutations originating from CX1834-1-6), it is degraded back into *myo*-inositol in the cytosol. This potential accumulation of *myo*-inositol could replace the *myo*-inositol that is not being produced due to the faulty MIPS1 gene in V99-

5089, and this *myo*-inositol could feed into the raffinosaccharide pathway. Further research into the interrelatedness of the phytate and raffinosaccharide pathways could shed light on these unexpected results.

INTRODUCTION

Sucrose, raffinose, and stachyose are the three main forms of carbohydrates found in soybean [*Glycine max* (L.) Merr.] seeds. The majority of seed oligosaccharides is sucrose, with levels usually ranging from 5-7%, stachyose at 3-4%, and raffinose at about 1% of seed dry weight (Skoneczka et al., 2009). Raffinose and stachyose are D-galactose-containing oligosaccharides which make up the raffinose family oligosaccharides (RFOs). These sugars consist of a chain of galactosyl residues attached by an Ŭ-(1→6) glycosidic bond to a sucrose molecule (Avigad and Dey, 1997). The majority of seed carbohydrate content is found as sucrose, with RFOs comprising the remaining fraction (Hymowitz and Collins, 1974). The main role of RFOs in plants is transporting photoassimilate from the leaves to the phloem (Ayre et al., 2003), or serving as storage carbohydrates and cryoprotectants to promote freezing-tolerance (Pennycooke et al., 2003; Sprenger and Keller, 2000). Seed RFOs are quickly broken down early during germination to provide the plant with energy and substrates needed for growth (Peterbauer and Richter, 2001). Perhaps the most widely known role of RFOs is in desiccation tolerance and longevity of seeds (Horbowicz and Obendorf, 1994).

RFO Biosynthesis

RFOs are synthesized by the reversible transfer of galactosyl units from galactinol or from other RFOs to a galactosyl acceptor, which may be sucrose or raffinose (Peterbauer and Richter, 2001). If the galactosyl acceptor is sucrose, raffinose synthase catalyzes the chain elongation by transferring galactosyl groups to the sucrose to make raffinose. Raffinose is then used by stachyose synthase to produce stachyose (Lehle and Tanner, 1973; Tanner and Kandler, 1968). However, before this takes place, galactinol synthase must synthesize galactinol from

UDP-D-galactose and *myo*-inositol (Keller and Pharr, 1996) (See Figure 4.1 for a simplified diagram).

Galactinol synthase has been regarded as a key enzyme in the RFO pathway, due to its product (galactinol) being required for RFO synthesis. In *Arabidopsis*, overexpression of galactinol synthase subsequently increases galactinol and raffinose (Taji et al., 2002). However, other studies suggest that not only does galactinol synthase act on RFO synthesis, but other metabolites such as *myo*-inositol act on RFO synthesis as well. For example, Peterbauer et al. (2001) found no significant difference in galactinol synthase expression and activity levels in two pea varieties with different RFO levels. Because *myo*-inositol is one of the substrates used by galactinol synthase, it is plausible that its concentrations affect RFO accumulation in seeds. Indeed, in the soybean mutant LR33, a reduction in *myo*-inositol lowered galactinol and raffinosaccharide levels significantly (Hitz et al., 2002). The enzyme *myo*-inositol phosphate synthase (MIPS) catalyzes the reaction of glucose-6-phosphate to *myo*-inositol-3-phosphate, which is a precursor to *myo*-inositol. MIPS is ubiquitous in diverse organisms, with the catalytic core domain being conserved (GhoshDastidar et al., 2006). Keller et al. (1998) used antisense RNA to suppress MIPS to a level of ~20% of the wild-type in transgenic potato, which resulted in drastically reduced levels of inositol, galactinol, and raffinose (Keller et al., 1998). Karner et al. (2004) found that higher seed *myo*-inositol levels were associated with an increase in galactinol content in pea seeds. However, this increase in galactinol did not necessarily promote RFO accumulation (Karner et al., 2004). Together, these results indicate that there is no key enzyme controlling the accumulation of RFOs, but rather a complex pathway depending on substrates such as *myo*-inositol, sucrose, and UDP-galactose.

Phytic Acid Biosynthesis and how it Relates to RFO Biosynthesis

Phytic acid (*myo*-inositol (1,2,3,4,5,6)-hexa*kis*phosphate) is the major form of phosphorus in seeds (Raboy, 2007). As mentioned previously, the first step in phytic acid biosynthesis proceeds through the conversion of glucose-6-phosphate to inositol-3-phosphate, catalyzed by MIPS. This is the first committed step in *myo*-inositol production, and is considered the sole synthetic source of *myo*-inositol (Loewus and Murthy, 2000). *Myo*-inositol is formed from the dephosphorylation of inositol-1-phosphate by inositol monophosphatase. Both the phytic acid and RFO pathways utilize free *myo*-inositol as a common intermediate, therefore phytate and RFOs are interrelated through their biochemical pathways.

Both phytic acid and RFO levels have become targets for manipulation in recent years in order to improve soybean seed quality (Wang et al., 2003). Because phytic acid chelates mineral cations including calcium, iron, and zinc, these mixed salts are often excreted by non-ruminant animals such as humans, swine, poultry, and fish (Brinch-Pedersen et al., 2002). In its salt form, phytic acid is known as phytate, and excretion of undigested phytate can cause water pollution due to the excess phosphorus (P) in animal waste that has the potential to run off into bodies of water (Sharpley et al., 1994). Various feeding trials have shown that low phytate grain can increase the available P to animals, satisfying more of their dietary requirement (Ertl et al., 1998; Jang et al., 2003). In turn, the animals excrete less P, resulting in less P runoff. In major cereal and legume crops, seeds normally contain between 3.0 and 8.0 mg of P per gram of dry weight, and out of this total P, 65-75% is phytic acid. About 5% of the remaining total P is inorganic phosphorus (Pi), and less than 10% is lower-order *myo*-inositol phosphates (Raboy, 2007). Raffinosaccharides are indigestible in monogastric animals and supply little metabolizable energy due to the fact that monogastric animals do not synthesize enough of the enzyme Ŭ-

galactosidase to digest these raffinosaccharides (Gitzelmann and Auricchio, 1965). Although raffinosaccharides can provide a beneficial probiotic function (Mussatto and Mancilha, 2007), they may also cause gastric discomfort, flatulence, and less metabolizable energy (Coon et al., 1990; Hata et al., 1991). In a field study where meal from low stachyose/raffinose soybean lines were fed to roosters, metabolizable energy was significantly higher than those roosters that was fed meal from normal stachyose and raffinose lines (Parsons et al., 2000). Therefore, sucrose is the more desirable seed sugar compared to the raffinosaccharides, which are considered anti-nutritional. Due to the indigestibility of phytate and RFOs in humans and animals and its environmental consequences, low phytate and raffinosaccharide varieties of crops are beneficial (Raboy, 2007).

Mutations Controlling Phytate or Oligosaccharide Content

Currently, most known low phytate and/or low RFO varieties of crops have been produced via mutagenesis, during which one or more of the genes involved in the phytate/raffinosaccharide pathway has been altered. These mutations presumably hinder the plantôs ability to synthesize or accumulate phytate or RFOs. Various genes have been shown to confer low phytate when mutated; for example inositol phosphate kinase, *myo*-inositol kinase, or *myo*-inositol phosphate synthase (MIPS) genes are known to cause low phytate (Pilu et al., 2003; Shi et al., 2005). Multidrug resistance-associated protein (MRP) ABC transporters have also been implicated as the cause for low phytate in several crops (Gillman et al., 2009; Panzeri et al., 2011; Saghai Maroof et al., 2009; Shi et al., 2007; Xu et al., 2009). The *Arabidopsis* gene AtMRP5 (closely related to the maize MRP4 gene), was shown to encode for an inositol hexa*kis*phosphate transporter, mutations in which can cause low phytate (Nagy et al., 2009). CX1834-1-6 (hereafter CX1834) is a low phytate soybean line, which was shown to have two

point mutations, one in an MRP gene on LG N (chromosome 3), and one in an MRP gene on LG L (chromosome 19) (Gillman et al., 2009; Saghai Maroof et al., 2009). These SNP-type mutations most likely disrupt the function of the proteins, and subsequently block phytate accumulation. Panzeri et al. (2011) hypothesized that since phytate cannot be properly stored in the vacuole, it gets broken down by phytases to its precursor, *myo*-inositol, in the cytoplasm. This could cause negative feedback regulation, leading to low phytate.

In addition to MRP genes, alteration of MIPS genes can lead to low phytate soybeans. Four MIPS homologs exist in soybean, designated as MIPS1, MIPS2, MIPS3, and MIPS4 (Chappell et al., 2006; Hegeman et al., 2001). LR33 is a low phytate mutant with about 50% lower phytate levels and reduced levels of raffinosaccharides (Sebastian et al., 2000). Hitz et al. (2002) measured levels of sucrose, galactinol, raffinose, and stachyose, and determined that the mutation in LR33 affects a step in the biochemical pathway that is needed for galactinol increase. In addition, *myo*-inositol was decreased by 36% in LR33 compared to wild type. *Myo*-inositol is known to be involved in the biosynthesis of raffinose and stachyose, as well as for being a precursor in the phytic acid pathway. These observations suggested decreased activity of MIPS, and upon sequencing the cDNA clone, it was determined that MIPS1 in LR33 has a mutation of a G to T (Hitz et al., 2002). The mutation causes an amino acid change from a lysine to a glutamine, and subsequently decreases the phytate and RFOs levels. It is unknown what structural effect results from this amino acid substitution, although the authors showed that this particular amino acid (lysine) is conserved in MIPS genes across several species including soybean, *Arabidopsis*, maize, *Drosophila*, and yeast (Hitz et al., 2002). Another study isolated and characterized a novel source of low phytate in soybean (Yuan et al., 2007), termed *Gm-lpa*-TW-1. *Gm-lpa*-TW-1 showed a reduction in seed RFOs as well as phytate, similar to the MIPS1

mutation in soybean mutant line LR33 (Hitz et al., 2002). The authors therefore hypothesized that low phytate in *Gm-lpa*-TW-1 was due to a mutation in MIPS1, and sequenced the gene to compare with wild-type. There was a 2 bp deletion in the coding region of MIPS1, causing a frame shift and most likely leading to the low phytate and low RFO phenotype (Yuan et al., 2007). A study by Nunes et al. (2006) used RNAi to silence the MIPS1 gene in soybean and observed up to a 94.5% decrease in phytic acid levels of transgenic seeds. Together, these studies demonstrate the integral role of MIPS genes in phytic acid biosynthesis.

V99-5089 is a Virginia Tech proprietary soybean line developed by traditional breeding methods, as opposed to mutagenesis. It has a low phytate phenotype, along with low RFO and high sucrose content, most likely due to a spontaneous mutation. Because of V99-5089ôs low phytate/stachyose phenotype, it is of great agronomic interest for possible use in breeding programs. Currently, there are only two other known soybean lines with both of these characteristics, LR33 and *Gm-lpa*-TW-1 (Hitz et al., 2002; Yuan et al., 2007; Yuan et al., 2009). Previous studies from our group have mapped the low phytate/low stachyose QTL to LG B1 (chromosome 11), and the map position of the low stachyose QTL corresponds to the *myo*-inositol 1-phosphate synthase (MIPS1) gene. Subsequent sequencing of MIPS1 shows that V99-5089 has a novel mutation that has not been previously reported. This function-affecting transversion is a single base change from a C to a G at position 1144 of the coding sequence of MIPS1 (Biyashev et al., manuscript in preparation).

In order to study the effects of combining both mutations from CX1834 (the SNP mutations in LG N and LG L on MRP genes) with the V99-5089 mutation on LG B1 (SNP mutation in the MIPS1 gene), we have developed a recombinant inbred population from a cross of V99-5089 x CX1834. This population has provided an opportunity to access all combinations

of the three mutations, and to observe the corresponding phytate and oligosaccharide phenotypes. The objective of this study was to determine the effects of different combinations of the MIPS (located on LG B1), the MRP (LG L), and the MRP (LG N) mutations on phytate and oligosaccharide content in the V99-5089 x CX1834 RIL population.

MATERIALS AND METHODS

Genetic Materials

The recombinant inbred population CX1834 x V99-5089 was used for marker and phenotypic data collection in this study. CX1834 is a low phytate line derived from a cross between óAthow,ô a normal phytate soybean cultivar, and M153-1-4-6-14, a low phytate mutant (Wilcox et al., 2000). V99-5089 is a Virginia Tech experimental line with low phytate, low RFOs, and high sucrose. The CX1834 x V99-5089 population has a total of 173 individuals that were used for collecting genotypic and phenotypic data. This population was planted at Kentland Farm, Virginia in 2006, 2007, 2008, 2009, and 2010 and seed was harvested for the F_6, F_7, F_8, F_9, and F_{10} generations, respectively. The soil at Kentland Farm is a Hayter loam (Gao et al., 2008). F_7 seeds were used to assay the whole population for phytate, and F_6 and F_8 seeds were used to assay the whole population for oligosaccharide content. F_8, F_9, and F_{10} seeds were used for the phytate assay on a subset of fourteen samples from the population (the mips/mrp-l/mrp-n genotypic class, see Results section), and F_{10} seeds were used for the oligosaccharide assay on the same subset of fourteen samples. Both parental lines (CX1834 and V99-5089) were planted in the field in 2008 along with the rest of the F_8 population. One bulk sample for each parent was collected and assayed for phytate and oligosaccharide content nine times for V99-5089 and seven times for CX1834.

Marker Data Collection

Approximately 30 F_8 V99-5089 x CX1834 seeds harvested from the field at Kentland Farm, Virginia were planted in the greenhouse at Virginia Tech in 2008. The seeds planted for each sample were a bulk of single plants, each bulk sample representing an RIL. Ten to sixteen plants for each RIL were sampled by collecting trifoliate leaf tissue for DNA extraction,

performed according to Yu et al. (1994). The locations of the MRP mutations in CX1834 on LGs L and N (Gillman et al., 2009; Saghai Maroof et al., 2009) as well as the MIPS mutation in V99-5089 (Biyashev et al., manuscript in preparation) were used to design primers for SNP genotyping. The RILs were genotyped for SNPs at the University of Georgia, using melting curve analysis with SimpleProbe (TIB MOLBIOL, Adelphia, NJ, USA). A set of primers was designed for the SNP detection were designated as SNP-L, SNP-N, and SNP-B1 for LGs L, N, and B1, respectively (n=173). DNA from the RILs from the CX1834 x V99-5089 population in the F_8 generation was used for marker data collection. These lines were genotyped for the SNP mutations on all three linkage groups. The SNP-L forward primer sequence is 5ô-CTGAATTTAAATGCACGTC-3ô, the reverse primer sequence is 5ô-TGTGAAGCTGAGGTTAG-3ô, and the SimpleProbe primer sequence was 5ó-TTGGCTGTACTGATA**XI**AATT**C**TCTCAATAGï Phosphate-3ô. The SNP-N forward primer sequence is 5ô-CCTGGAGGCATCTGTTATGAC-3ô, the reverse primer sequence is 5ô-CTGCCATGTATGAAAGAT-3ô, and the SimpleProbe primer sequence was 5ô-CAAGCTGTT**XI**TCTTTC**A**CGATCGTT--Phosphate-3ô. The SNP-B1 forward primer is sequence is 5ô-AACAATGATGGTATGAATCTTTCG-3ô, the reverse primer sequence is 5ô-CCTGACAAGAGAAAGAAACAGA-3,ó and the SimpleProbe primer sequence was 5ó-gTgAACA**XI**TCCAgAC**C**ATgTTgTTgTT-Phosphate-3.ôñXIò is the internal SimpleProbe label, and the position of the SNP in the probes is underlined and bolded. Fluorescein was used to internally label the SimpleProbes. A LightCycler 480 (Roche Applied Science, Indianapolis, IN, USA) was used for PCR reactions, with a total volume of 3 µL per well, using the protocol for asymmetrical PCR. Each PCR reaction consisted of 20-30 ng of genomic DNA, 0.5 µM of

limiting primer, 1.0 μM of excess primer, 0.2 μM of SimpleProbe, 3.0 mM $MgCl_2$, and 0.5X of LightCycler 480 Genotyping Master mix (Roche Applied Science, Indianapolis, IN, USA).

Fifty PCR cycles were performed with 10 s of denaturation at 95°C, 15 s of annealing at 55°C, and 20 s extension at 72°C. A final melting cycle was performed at 95°C for 2 min, 40°C for 2 min, and increasing the temperature to 85°C, at this point using continuous fluorescent acquisition, and followed by a decrease in temperature to 40°C. To generate melting curves for each sample, the fluorescence signal (*F*) was plotted against temperature (*T*) in real time. Negative derivative curves of fluorescence with respect to temperature (-d*F*/d*T*) were produced by the LightCycler Data Analysis software (Roche Diagnostics, Indianapolis, IN, USA).

In order to confirm the SNP data for the subset of fourteen mips/mrp-l/mrp-n samples from this population, the samples were sequenced for the LG L, N, and B1 mutations with the procedure according to Saghai Maroof et al. (2009).

Phytate Data Collection

Seeds from the F_7, F_8, F_9, and F_{10} generations from the CX1834 x V99-5089 population in the field at Kentland Farm were harvested in 2007, 2008, 2009, and 2010. Ground powder from approximately 75 seeds from each of the 173 RILs from the F_7 generation were used in the phytate assay, conducted using a modified colorimetric method (Gao et al., 2007). The phytate assay was only conducted on a subset of the lines using F_8, F_9, and F_{10} generation seed. Briefly, 0.5 g of soybean powder was weighed into 14 mL falcon tubes. To each tube, 10 mL of 0.65 M HCl was added, and the tubes were vortexed and put on a shaker overnight (220 rpm at room temperature). The samples were then centrifuged at 3300 rpm at 10°C for 15 min, and 500 μL of the supernatant was transferred to a microcentrifuge tube and mixed with 500 μL of 20% NaCl solution. After a 2 hr precipitation of the crude extract, the samples were centrifuged at 13,200

rpm for 15 min. From each sample, 120 µL of the supernatant was added to 2.88 mL of ddH_2O (25 times dilution), and an additional 1 mL of Wadeôs Reagent (0.03% $FeCl_3 \cdot 6H_2O$ + 0.3% sulfosalicylic acid) was added for color development. Next, the samples were centrifuged at 3500 rpm for 10 min at 10°C, and after centrifugation, the samples were read on a Beckman Coulter DU 800 Spectrophotometer (Fullerton, CA) at 500 nm to determine phytate content. Phytate concentrations of the samples were calculated using a calibration curve consisting of eight standards (0, 1.12, 2.24, 3.36, 5.6, 7.84, 8.96, and 11.2 ppm phytate). These standards also contained HCl and NaCl to decrease matrix effects.

Oligosaccharide Data Collection

Oligosaccharide data was collected for the whole population using F_6, and F_8 generations of seed. F_{10} seeds were used only for the subset of fourteen samples. Determination of sugar content in soybean seeds by high performance liquid chromatography (HPLC) was based on the procedure originally described by Cicek et al. (2006), with modifications. Briefly, about 1.000 g of ground sample from each of the 173 RILs in the population was weighed into 12 mL centrifuge tubes, and 10 mL of ddH_2O were added to each sample. The samples were mixed by vortexing and shaking at 200 rpm for 20 min. A centrifugation step was next, for 10 min at 4000 rpm. From these centrifuged samples, 0.5 mL of supernatant was transferred into another 1.5 mL centrifuge tube, and 0.7 mL of 100% acetonitrile was added to each tube. The contents of the tube were mixed and allowed to sit at room temperature for 2 hrs. Next, 1.0 mL of the supernatant was taken from each sample and was dried at 80°C under air flow. The samples were then dissolved with 0.5 mL of 65% acetonitrile, centrifuged at 13,000 rpm for 10 min, and transferred to HPLC vials for analysis.

Sucrose, raffinose, and stachyose were measured with an Agilent 1200 high performance liquid chromatograph with a differential refractrometer detector (refractive index detector) (Santa Clara, CA). A Supelco apHera NH_2 analytical column (4.6 x 250 mm, 5 µm) was used for separation, along with a Supelco apHera NH_2 guard column (1 cm x 4.6 mm, 5 µm). All reagents were prepared using ddH_2O, and chemicals were all analytical or HPLC grade. The mobile phase, 65% acetonitrile, was made by combining 65 mL of 100% acetonitrile and 35 mL of ddH_2O and filtering through a 0.45 µm filter. The elution program used was: 65% ACN with a flow rate of 1 mL min^{-1}, and 10 µL from each sample was used for injection.

Statistical Analysis

Since we were only interested in the homozygous lines, recombinant inbred lines that were heterozygous at either of the LG L, N, or B1 mutation sites were removed from the data set. All statistical analyses including descriptive statistics and frequency distributions were performed with JMP v.8 software (SAS Institute Inc., Cary, NC). Pairwise comparison of means was based on Tukey-Kramerôs Honestly Significant Difference (HSD) and Studentôs t least squared mean tests at the Ŭ= 0.05 level.

RESULTS AND DISCUSSION

SNP Genotyping

Three known mutations control oligosaccharide content and/or phytate content in the V99-5089 x CX1834 population: soybean line V99-5089 has a SNP-type mutation on LG B1 with a ñCò as the wild-type and ñGò as the mutant. This mutation is in the MIPS1 gene, and previous studies have shown that this particular mutation is the cause of low stachyose/high sucrose/low phytate in V99-5089 (Biyashev et al., manuscript in preparation). CX1834 has two low phytate-causing mutations: an ñAò to a ñTò mutation on LG N, and a ñGò to ñAò mutation on LG L. The latter two mutations occur in two different MRP ABC transporter genes and have been shown to be responsible for low phytate in CX1834 (Gillman et al., 2009; Saghai Maroof et al., 2009).

SNP data for the genotypes at the low phytate/low stachyose mutation sites on LGs B1, L and N were collected for 173 RILs of the V99-5089 x CX1834 population in the F_8 generation. Of those 173 RILs, ten were dropped because they were heterozygous at one or more mutation sites. The remaining 163 RILs could be divided into eight genotypic classes, each representing different combinations of the mutations on LGs B1, L, and N. The genotypic classes are as follows, represented as ñgenotype at B1 mutation site/genotype at L mutation site/genotype at N mutation siteò: mips/MRP-L/MRP-N, mips/MRP-L/mrp-n, mips/mrp-l/MRP-N, mips/mrp-l/mrp-n, MIPS/mrp-l/mrp-n, MIPS/MRP-L/mrp-n, MIPS/mrp-l/MRP-N, MIPS/MRP-L/MRP-N. Lower case letters represent the mutant allele, and uppercase letters represent the wild-type allele for the B1 MIPS, L MRP, and N MRP loci, respectively. For example, the ñmips/MRP-L/MRP-Nò genotypic class has only the B1 MIPS mutation (i.e., V99-5089 allele at all three mutation sites; parental type 1). The genotypic class represented by ñMIPS/mrp-l/mrp-nò contains the L

and N MRP mutations (i.e., CX1834 alleles at all three mutation sites; parental type 2). (For a quick reference for the definitions of each class, see Supplementary Table 4.1).

Assuming no unintended selection occurred due to lowered emergence of the low phytate seeds, we should observe a 1:1:1:1:1:1:1:1 segregation ratio for these genotypic classes. The χ^2 value was 14.32 with 7 degrees of freedom and a two-tailed p-value of 0.045. This χ^2 indicates that the difference between the observed and expected values was statistically significant. (A p-value less than 0.05 indicates that the data do not follow the expected distribution.) The numbers of RILs in each of the eight genotypic classes are shown in Table 4.1. The genotypic classes with the least number of RILs per class (14 lines each) had all three mutations (mips/mrp-l/mrp-n), or the B1 and N mutation but not the L mutation (mips/MRP-L/mrp-n). The genotypic classes with the most individuals (29 lines) were MIPS/MRP-L/MRP-N and MIPS/mrp-l/MRP-N, which are wild type at the LG B1 mutation site, and wild type at the LG N and/or the LG L mutation site.

Phytate Content

In soybean seeds, phytate content normally ranges from about 5-18 mg g^{-1} (Gao et al., 2007). The phytate values for the parents of this population were as follows: V99-5089 at 10.62 mg g^{-1} and CX1834 at 8.78 mg g^{-1} (Table 4.1); both parents are considered low phytate lines. The modified colorimetric phytate assay (Gao et al., 2008; Gao et al., 2007) was used to directly measure the phytate content of all of the V99-5089 x CX1834 RILs of the F_7 generation. The phytate values for the RILs ranged from 4.01 to 17.83 mg g^{-1}, and the mean was 11.81 mg g^{-1}. When considering the frequency distribution of the phytate values for the RILs in this population (Figure 4.2), there appears to be no normality over the population as a whole, but rather three smaller peaks that appear normal: one from approximately 5-6 mg g^{-1}, one from 9-10 mg g^{-1}, and

one from 14-15 mg g^{-1}. These three clusters of phytate values most likely represent the phytate values of the different combinations of mutationsð those with the CX1834 mutations on LGs L and N from 5-6 mg g^{-1} (lowest phytate), those with the V99-5089 mutation on B1 from 9-10 mg g^{-1} (low phytate, but not as low as CX1834), and those with wild type alleles from 14-15 mg g^{-1} (high phytate).

Mean phytate values for each of the genotypic classes combining the various mutations can be seen in Table 4.1. The mean phytate values for each class appear to fall into the same three categories as seen in the frequency distributions. The genotypic classes with the lowest phytate (mips/mrp-l/mrp-n and MIPS/mrp-l/mrp-n), have either all three low phytate mutations or only the LGs L and N mutations. The phytate for these two classes are 6.72 and 6.94 mg g^{-1} for mips/mrp-l/mrp-n and MIPS/mrp-l/mrp-n, respectively. This indicates that only the LGs L and N mutations are needed to achieve the lowest levels of phytate. The mips/mrp-l/mrp-n class also has one of the fewest numbers of RILs. This segregation distortion could be due to the reported poor emergence of low phytate seeds (Hulke et al., 2004; Oltmans et al., 2004). The next grouping of phytate values is midrange, around 10 mg g^{-1}, and consists of genotypic classes mips/MRP-L/MRP-N, mips/MRP-L/mrp-n, and mips/mrp-l/MRP-N. These all have the LG B1 mutation, combined with either LG L, LG N, or neither mutation. The phytate values of these classes indicate that combining only one of the CX1834 mutations with the V99-5089 mutation on LG B1 does not significantly lower phytate any lower than that of V99-5089. Those with the highest phytate (around 14 mg g^{-1}) were MIPS/MRP-L/MRP-N, MIPS/MRP-L/mrp-n, and MIPS/mrp-l/MRP-N. All have the wild type allele at LG B1, and then a wild type allele at either LG L, LG N, or both mutation sites. Based on this data set, one can achieve the lowest phytate content by combining all three mutant alleles, or alternatively, by combining just the LG L and

LG N mutations. The L and N mutations together will result in just as low phytate as combining all three mutations, as the values for the mips/mrp-l/mrp-n and MIPS/mrp-l/mrp-n classes were not significantly different. The next lowest level of phytate can be achieved by simply having the LG B1 mutation combined with either of the L and N mutations, or just by the LG B1 mutation alone with wild-type L and N alleles. From a practical standpoint, it would be worthwhile to compare the average emergence for the mips/mrp-l/mrp-n and MIPS/mrp-l/mrp-n classes. If one class has significantly greater emergence, it would be the better choice for a breeding program, as one of the main difficulties to overcome with low phytate crops is the lowered seedling emergence (Hulke et al., 2004; Oltmans et al., 2004).

Oligosaccharide Content

The two parents of this population have contrasting oligosaccharide values, especially sucrose and stachyose. V99-5089 is a high sucrose, low stachyose line with a mean sucrose at 12.5% and stachyose at 0.25% of the dry seed weight (n=9 assays). CX1834 has a low sucrose value, averaging 7.14%, and high stachyose, at 4.48% (n=7 assays). Oligosaccharide data were collected for the whole population as well. F_8 generation seed was assayed for the 173 RILs. For sucrose, the mean for all the RILs in the population was 8.29%, and ranged from 4.52 - 13.65%. The raffinose mean was 0.73%, with a range from 0.18 ï 1.71 %. Stachyose averaged 3.04%, and ranged from 0.13 ï 5.52%. Because of the wide range of sucrose and stachyose values and the existence of transgressive segregants, there is the possibility to breed for lines with even lower levels of stachyose and higher levels of sucrose than V99-5089. The frequency distributions for sucrose, stachyose, and raffinose are not normal, but appear to have a bimodal distribution, especially stachyose (Figure 4.3).

We then examined the sugar data according to genotypic class (Table 4.1). Regardless of the L or N MRP alleles, all the classes with the V99-5089 mutant allele at the LG B1 mutation site (mips) had significantly greater mean sucrose level and three of these four classes had significantly lower mean raffinose and stachyose compared to those with the CX1834 allele at the B1 mutation site. All classes with the CX1834 allele (wild-type) at the B1 mutation site (MIPS) had sucrose values at about 7%, and high raffinose and stachyose, over 0.79% and 3.46% respectively (Table 4.1). This is expected considering the L and N mutations are only known to affect phytate and not sugar content. The mips/MRP-L/mrp-n class (with the B1 mutation, wild type L, N mutation) had the highest sucrose, at 11.16%. This class also had the second lowest raffinose value (0.42%) and the lowest stachyose value (0.54%). The mips/mrp-l/MRP-N class (B1 mutation, L mutation, wild type N) had similar sugar values to that of the mips/MRP-L/mrp-n class. It appears that the ñbestò lines (with the highest sucrose and lowest stachyose) can be achieved by combining the B1 mutation with either the L or the N mutations. The mips/MRP-L/MRP-N genotypic class, with only the MIPS mutation (wild type at the L and N mutation sites) had relatively high sucrose (9.67%), and low raffinose and stachyose (0.49% and 1.46%, respectively) (Table 4.1). It is unclear as to why the stachyose value is not as low as mips/MRP-L/mrp-n or mips/mrp-l/MRP-N classes, which are both closer to the V99-5089 parental value (0.25%).

ñmips/mrp-l/mrp-nò Genotypic Class

The oligosaccharide values for the mips/mrp-l/mrp-n class were unexpected (Table 4.1). This genotypic class has all three mutations, and one would have expected this class to be low phytate as well as high sucrose and low stachyose (low phytate being conferred by V99-5089 and CX1834, and low stachyose being conferred by V99-5089). However, although this class

was low phytate, it was neither high sucrose nor low stachyose (Table 4.1). Due to these unexpected results of the mips/mrp-l/mrp-n class, we collected several more sets of phenotypic and genotypic data in later generations in order to confirm the results (Table 4.2). We first confirmed the SNP marker data for the 14 mips/mrp-l/mrp-n individuals by sequencing the B1, L, and N mutation sites. The sequencing results confirmed the SNP data, that these lines did in fact have all three mutations. We then collected phytate data for the F_8, F_9, and F_{10} generations to compare with the F_7 phytate data we had previously obtained. All the 14 mips/mrp-l/mrp-n lines were confirmed to be low phytate. Next, we collected sugar data for the F_6, F_8, and F_{10} generations as well. Again the results were confirmed, and all the mips/mrp-l/mrp-n lines were high stachyose, instead of the low stachyose that one would expect with the MIPS mutation. In is unclear as to why we observed these high stachyose values whenever all three mutations are combined.

These unexpected results could possibly be explained by examining the interrelatedness of the phytate and raffinosaccharide pathways. MRP ABC transporters affect the accumulation of phytate in seeds. We know that the LG L and LG N mutations in CX1834 affect the function of two MRP ABC transporters, and that both mutations must be present to result in low phytate seeds. In *Arabidopsis*, MRP ABC transporters have been shown to control anion channels across the plasma membranes of guard cells and to confer partial drought tolerance (Suh et al., 2007). The gene controlling these traits, *AtMRP5*, was shown to encode an ATP-dependent inositol hexa*kis*phosphate transporter, and insertional mutants for this gene displayed the low phytate phenotype (Nagy et al., 2009). ABC transporters have also been shown to be responsible for low phytate in certain maize, rice, and common bean mutants (Panzeri et al., 2011; Shi et al., 2007; Xu et al., 2009). It is believed that the ABC transporters move phytate from the cytoplasm

across the membrane of the protein storage vacuoles (PSVs), to be used later by germinating seeds. The mutant transporters presumably block phytate from the PSVs, leaving free phytate in the cytoplasm (Raboy, 2009). We hypothesized that this phytate in the cytoplasm could be broken down into its precursors, free *myo*-inositol and lower-order phosphates.

The enzyme MIPS converts glucose-6-phosphate to inositol-3-phosphate, which is a precursor to *myo*-inositol. In V99-5089, the mutation in MIPS1 most likely disrupts this function, subsequently causing no *myo*-inositol to be produced. Since m*yo*-inositol is a precursor to raffinosaccharides, a mutation in MIPS1 that disrupts its function ultimately leads to lower raffinosaccharidesð because *myo*-inositol cannot be synthesized. However, if free *myo*-inositol is present in the cytoplasm due to the inability of the MRP ABC transporters to transport phytate into the vacuoles, we think that this *myo*-inositol in the cytoplasm could serve as the precursor to RFOs, even with a faulty MIPS1. Hence, this could explain why the mips/mrp-l/mrp-n class has high stachyose and low sucrose, when the opposite is expected (Figure 4.3).

This hypothesis is not in agreement with the theory of Panzeri et al (2011). In their study, they used the low phytate *Phaseolus* mutant line *lpa1*(280-10), and found a 90% reduction in phytate and a 25% reduction in raffinosaccharides (Campion et al., 2009). A single base mutation in an MRP gene was shown to cause the mutant phenotype, and no accumulation of phytate at any stage of seed development was observed in the mutant line, although it could still synthesize phytate (Panzeri et al., 2011). These results, along with the observation of less *myo*-inositol in the common bean mutant line 280-10, suggested that the MRP mutation may reduce phytate biosynthesis through negative feedback. According to Panzeri et al. (2011), because phytate cannot be transported to the vacuoles, it is degraded back to *myo*-inositol by phytases. This proposed negative feedback may decrease *myo*-inositol and subsequently decrease RFOs,

since *myo*-inositol feeds into the RFO pathway. This is in disagreement with our suggested model, that a faulty MRP would increase *myo*-inositol accumulation in the cytosol and subsequently increase RFOs. More work studying the biochemical mechanism controlling *myo*-inositol, phytate, and other intermediates should be done to shed light on the reason as to why the mutant MRP in *lpa1*(280-10) cause low phytate AND reduced RFOs, whereas CX1834ôs mutant MRP only causes low phytate.

This study used three SNP markers, one for each mutant allele, to easily genotype a recombinant inbred population. These SNP markers could be useful in breeding programs for marker-assisted selection of phytate and oligosaccharide content. Since SNPs are the causative mutation for the mutant phenotype in all three genes, these markers are perfect markers and should be more effective than phenotypic selection, or even selection using tightly linked SSRs.

CONCLUSIONS

The V99-5089 x CX1834 population provided us with the unique opportunity to study the effects of three mutations on phytate and oligosaccharide levels on a population-wide level. The two MRP mutations on LGs L and N confer low phytate in soybean line CX1834. The mutation on the MIPS1 gene (LG B1) coming from V99-5089 confers not only low phytate, but low stachyose as well, making it useful in breeding programs. Upon examining all phytate and sugar data broken down by genotypic class, we observed a statistically significant segregation distortion among the eight different genotypic classes, most likely due to reduced emergence of those individuals with low phytate mutations. The genotypic classes with the lowest phytate were MIPS/mrp-l/mrp-n (L and N mutations, wild type B1), and mips/mrp-l/mrp-n (all three mutations). This indicated that only the L and N mutations are needed to achieve low phytate and that combining all three mutations does not necessarily mean even lower phytate than the parents. However, when looking at oligosaccharide content, only the B1 mutation confers low stachyoseð except in the case of the mips/mrp-l/mrp-n genotypic class. Although this class had the B1 mutation, combining it with the L and N mutations caused low phytate and high stachyose instead of low phytate and low stachyose. The reasoning as to why this happens is unclear and needs to be further investigated.

REFERENCES

Avigad G., Dey P.M. (1997) Carbohydrate metabolism: Storage carbohydrates, in: P. M. Dey, Harbourne, J. (Ed.), Plant Biochemistry, Academic Press, San Diego. pp. 143-204.

Ayre B.G., Keller F., Turgeon R. (2003) Symplastic continuity between companion cells and the translocation stream: long-distance transport is controlled by retention and retrieval mechanisms in the phloem. Plant Physiol. 131:1518-1528.

Brinch-Pedersen H., Sørensen L.D., Holm P.B. (2002) Engineering crop plants: getting a handle on phosphate. Trends in Plant Science 7:118-125.

Campion B., Sparvoli F., Doria E., Tagliabue G., Galasso I., Fileppi M., Bollini R., Nielsen E. (2009) Isolation and characterisation of an lpa (low phytic acid) mutant in common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. 118:1211-1221.

Chappell A.S., Scaboo A.M., Wu X., Nguyen H., Pantalone V.R., Bilyeu K.D. (2006) Characterization of the MIPS gene family in *Glycine max*. Plant Breeding 125:493-500.

Cicek M.S., Chen P., Saghai Maroof M.A., Buss G.R. (2006) Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. Crop Sci. 46:1253-1259.

Coon C.N., Leske K.L., Akavanichan O., Cheng T.K. (1990) Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. Poultry Science 69:787-793.

Ertl D.S., Young K.A., Raboy V. (1998) Plant Genetic Approaches to Phosphorus Management in Agricultural Production. J Environ Qual 27:299-304.

Gao Y., Biyashev R.M., Saghai Maroof M.A., Glover N.M., Tucker D.M., Buss G.R. (2008) Validation of low phytate QTLs and evaluation of seedling emergence of low phytate soybeans. Crop Sci. 48:1355-1364.

Gao Y., Shang C., Saghai Maroof M.A., Biyashev R.M., Grabau E.A., Kwanyuen P., Burton J.W., Buss G.R. (2007) A modified colorimetric method for phytic acid analysis in soybean. Crop Sci. 47:1797-1803.

GhoshDastidar K., Chatterjee A., Chatterjee A., Majumder A.L. (2006) Evolutionary divergence of L-*myo*-inositol 1-phosphate synthase: significance of a "core catalytic structure", in: A. L. Majumder, Biswas, B.B (Ed.), Biology of Inositols and Phosphoinositides. pp. 313-338.

Gillman J.D., Pantalone V.R., Bilyeu K. (2009) The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. The Plant Genome 2:179-190.

Gitzelmann R., Auricchio S. (1965) The handling of soya alpha-galactosides by a normal and galactosemic child. Pediatrics 36:231-235.

Hata Y., Yamamoto M., Nakajima K. (1991) Effects of soybean oligosaccharides on human digestive organs-- estimation of 50-percent effective dose and maximum noneffective dose based on diarrhea. Journal of Clinical Biochemistry and Nutrition 10:135-144.

Hegeman C.E., Good L.L., Grabau E.A. (2001) Expression of D-*myo*-inositol-3-phosphate synthase in soybean. Implications for phytic acid biosynthesis. Plant Physiol. 125:1941-1948.

Hitz W.D., Carlson T.J., Kerr P.S., Sebastian S.A. (2002) Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiol. 128:650-660.

Horbowicz M., Obendorf R.L. (1994) Seed desiccation tolerance and storability: Dependence on flatulence-producing oligosaccharides and cyclitols-- review and survey. Seed Science Research 4:385-405.

Hulke B.S., Fehr W.R., Welke G.A. (2004) Agronomic and seed characteristics of soybean with reduced phytate and palmitate. Crop Sci. 44:2027-2031.

Hymowitz T., Collins F.I. (1974) Variability of sugar content in seed of *Glycine max* (L.) Merrill and *G. soja* Sieb. and Zucc.1. Agron. J. 66:239-240.

Jang D., Fadel J., Klasing K., Mireles A., Jr, Ernst R., Young K., Cook A., Raboy V. (2003) Evaluation of low-phytate corn and barley on broiler chick performance. Poult Sci 82:1914-1924.

Karner U., Peterbauer T., Raboy V., Jones D.A., Hedley C.L., Richter A. (2004) *myo*-inositol and sucrose concentrations affect the accumulation of raffinose family oligosaccharides in seeds. Journal of Experimental Botany 55:1981-1987.

Keller F., Pharr D.M. (1996) Metabolism of carbohydrates in sinks and sources: galactosyl-sucrose oligosaccharides, in: E. Zamski, Schaffer, A.A. (Ed.), Photoassimilate distribution in plants and crops, Marcel Dekker, New York. pp. 115-184.

Keller R., Brearley C.A., Trethewey R.N., Müller-Röber B. (1998) Reduced inositol content and altered morphology in transgenic potato plants inhibited for 1D-*myo*-inositol 3-phosphate synthase. The Plant Journal 16:403-410.

Lehle L., Tanner W. (1973) The function of *myo*-inositol in the biosynthesis of raffinose. European Journal of Biochemistry 38:103-110.

Loewus F.A., Murthy P.P.N. (2000) *Myo*-inositol metabolism in plants. Plant Science 150:1-19.

Mussatto S.I., Mancilha I.M. (2007) Non-digestible oligosaccharides: A review. Carbohydrate Polymers 68:587-597.

Nagy R., Grob H., Weder B., Green P., Klein M., Frelet-Barrand A., Schjoerring J.K., Brearley C., Martinoia E. (2009) The *Arabidopsis* ATP-binding cassette protein AtMRP5/AtABCC5 is a high affinity inositol hexa*kis*phosphate transporter involved in guard cell signaling and phytate storage. Journal of Biological Chemistry 284:33614-33622.

Nunes A., Vianna G., Cuneo F., Amaya-Farfán J., de Capdeville G., Rech E., Aragão F. (2006) RNAi-mediated silencing of the *myo*-inositol-1-phosphate synthase gene (*Gm*MIPS1) in transgenic soybean inhibited seed development and reduced phytate content. Planta 224:125-132.

Oltmans S.E., Fehr W.R., Welke G.A., Cianzio S.R. (2004) Inheritance of low-phytate phosphorus in soybean. Crop Sci. 44:433-435.

Panzeri D., Cassani E., Doria E., Tagliabue G., Forti L., Campion B., Bollini R., Brearley C.A., Pilu R., Nielsen E., Sparvoli F. (2011) A defective ABC transporter of the MRP family, responsible for the bean *lpa1* mutation, affects the regulation of the phytic acid pathway, reduces seed *myo*-inositol and alters ABA sensitivity. New Phytologist:70-83.

Parsons C., Zhang Y., Araba M. (2000) Nutritional evaluation of soybean meals varying in oligosaccharide content. Poult Sci. 79:1127-1131.

Pennycooke J.C., Jones M.L., Stushnoff C. (2003) Down-regulating α-galactosidase enhances freezing tolerance in transgenic petunia. Plant Physiol. 133:901-909.

Peterbauer T., Richter A. (2001) Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. Seed Science Research 11:185-197.

Peterbauer T., Lahuta L.B., Blochl A., Mucha J., Jones D.A., Hedley C.L., Gorecki R.J., Richter A. (2001) Analysis of the raffinose family oligosaccharide pathway in pea seeds with contrasting carbohydrate composition. Plant Physiol. 127:1764-1772.

Pilu R., Panzeri D., Gavazzi G., Rasmussen S.K., Consonni G., Nielsen E. (2003) Phenotypic, genetic and molecular characterization of a maize low phytic acid mutant (lpa241). Theor. Appl. Genet. 107:980-7.

Raboy V. (2007) Seed phosphorus and the development of low-phytate crops, in: B. L. Turner, Richardson, A.E., and Mullaney, E.J. (Ed.), Inositol Phosphates: Linking Agriculture and the Environment, CAB International, Oxfordshire, UK. pp. 111-132.

Raboy V. (2009) Approaches and challenges to engineering seed phytate and total phosphorus. Plant Science 177:281-296.

Saghai Maroof M.A., Glover N.M., Biyashev R.M., Buss G.R., Grabau E.A. (2009) Genetic basis of the low-phytate trait in the soybean line CX1834. Crop Sci. 49:69-76.

Sebastian S.A., Kerr P.S., Pearlstein R.W., Hitz W.D. (2000) Soybean germplasm with novel genes for improved digestibility, in: J. K. Drackley (Ed.), Soy in Animal Nutrition, Federation of Animal Science Societies, Savoy, IL. pp. 56-74.

Sharpley A.N., Chapra S.C., Wedepohl R., Sims J.T., Daniel T.C., Reddy K.R. (1994) Managing agricultural phosphorus for protection of surface waters - issues and options. Journal of Environmental Quality 23:437-451.

Shi J., Wang H., Hazebroek J., Ertl D.S., Harp T. (2005) The maize low-phytic acid 3 encodes a *myo*-inositol kinase that plays a role in phytic acid biosynthesis in developing seeds. The Plant Journal 42:708-719.

Shi J., Wang H., Schellin K., Li B., Faller M., Stoop J.M., Meeley R.B., Ertl D.S., Ranch J.P., Glassman K. (2007) Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. Nat. Biotech 25:930-937.

Skoneczka J.A., Saghai Maroof M.A., Shang C., Buss G.R. (2009) Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. Crop Sci. 49:247-255.

Sprenger N., Keller F. (2000) Allocation of raffinose family oligosaccharides to transport and storage pools in Ajuga reptans: the roles of two distinct galactinol synthases. The Plant Journal 21:249-258.

Suh S.J., Wang Y.-F., Frelet A., Leonhardt N., Klein M., Forestier C., Mueller-Roeber B., Cho M.H., Martinoia E., Schroeder J.I. (2007) The ATP binding cassette transporter AtMRP5 modulates anion and calcium channel activities in *Arabidopsis* guard cells. Journal of Biological Chemistry 282:1916-1924.

Taji T., Ohsumi C., Iuchi S., Seki M., Kasuga M., Kobayashi M., Yamaguchi-Shinozaki K., Shinozaki K. (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. The Plant Journal 29:417-426.

Tanner W., Kandler O. (1968) *Myo*-Inositol, a cofactor in the biosynthesis of stachyose. European Journal of Biochemistry 4:233-239.

Wang T.L., Domoney C., Hedley C.L., Casey R., Grusak M.A. (2003) Can we improve the nutritional quality of legume seeds? Plant Physiol. 131:886-891.

Wilcox J.R., Premachandra G.S., Young K.A., Raboy V. (2000) Isolation of high seed inorganic P, low-phytate soybean mutants. Crop Sci. 40:1601-1605.

Xu X.-H., Zhao H.-J., Liu Q.-L., Frank T., Engel K.-H., An G., Shu Q.-Y. (2009) Mutations of the multi-drug resistance-associated protein ABC transporter gene 5 result in reduction of phytic acid in rice seeds. Theor. Appl. Genet. 119:75-83.

Yu Y.G., Saghai Maroof M.A., Buss G.R., Maughan P.J., Tolin S.A. (1994) RFLP and microsatellite mapping of gene for soybean mosaic virus resistance. Phytopathology 84:60-64.

Yuan F.-J., Zhao H.-J., Ren X.-L., Zhu S.-L., Fu X.-J., Shu Q.-Y. (2007) Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). Theor. Appl. Genet. 115:945-957.

Yuan F.-J., Zhu D.-H., Deng B., Fu X.-J., Dong D.-K., Zhu S.-L., Li B.-Q., Shu Q.-Y. (2009) Effects of two low phytic acid mutations on seed quality and nutritional traits in soybean (*Glycine max* L. Merr). Journal of Agricultural and Food Chemistry 57:3632-3638.

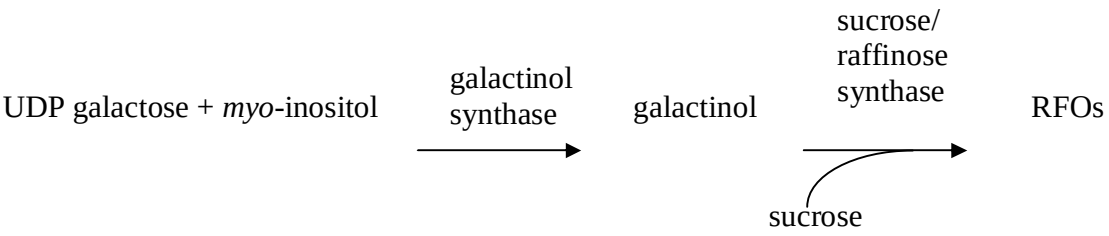


Figure 4.1. Schematic diagram of the raffinosaccharide (RFO) pathway.

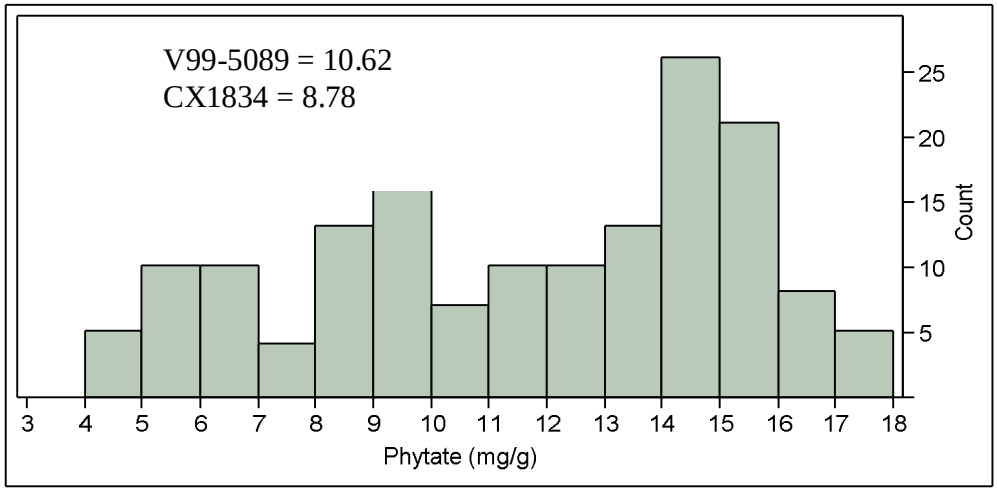


Figure 4.2. Frequency distribution of the phytate trait in 163 lines derived from the V99-5089 x CX1834 recombinant inbred population, at the F_7 generation. Phytate values for each parent are displayed on the graph.

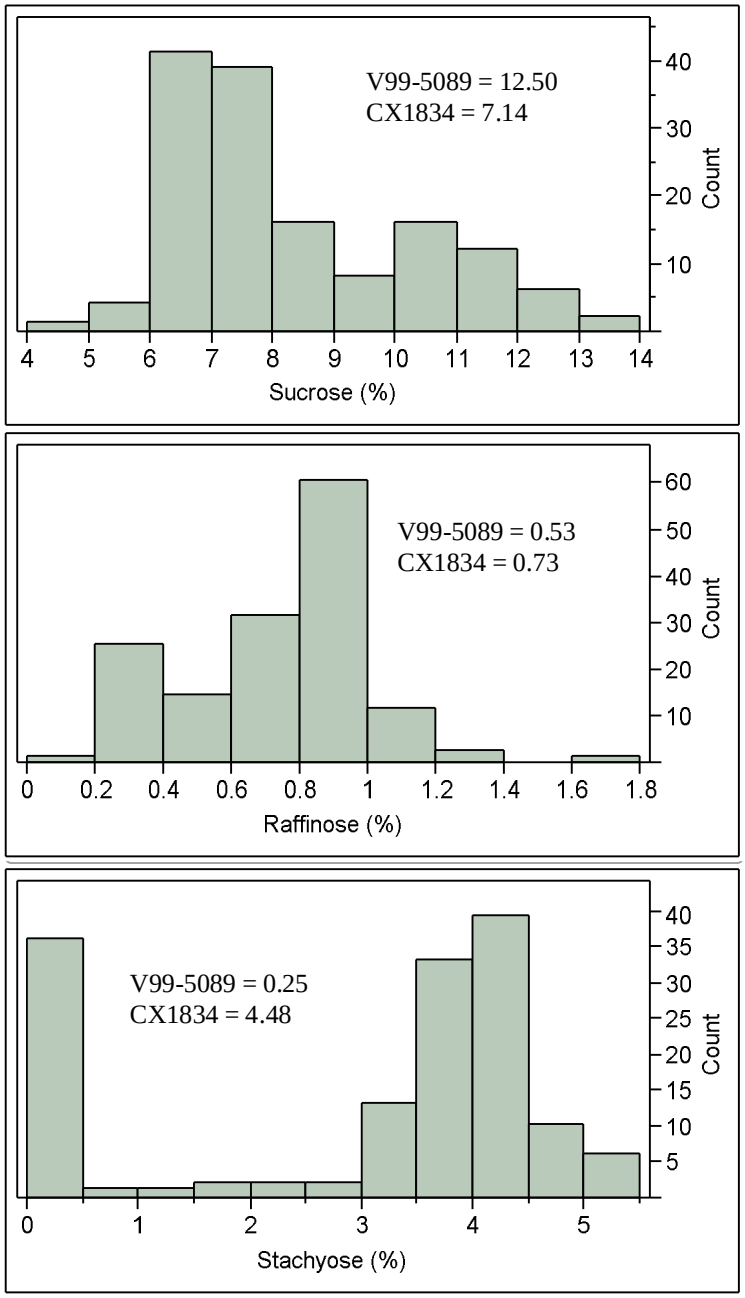


Figure 4.3. Frequency distributions of the % sucrose, raffinose, and stachyose trait in 163 lines derived from the V99-5089 x CX1834 recombinant inbred population, at the F_6 generation. Oligosaccharide values for each parent are displayed on the graph.

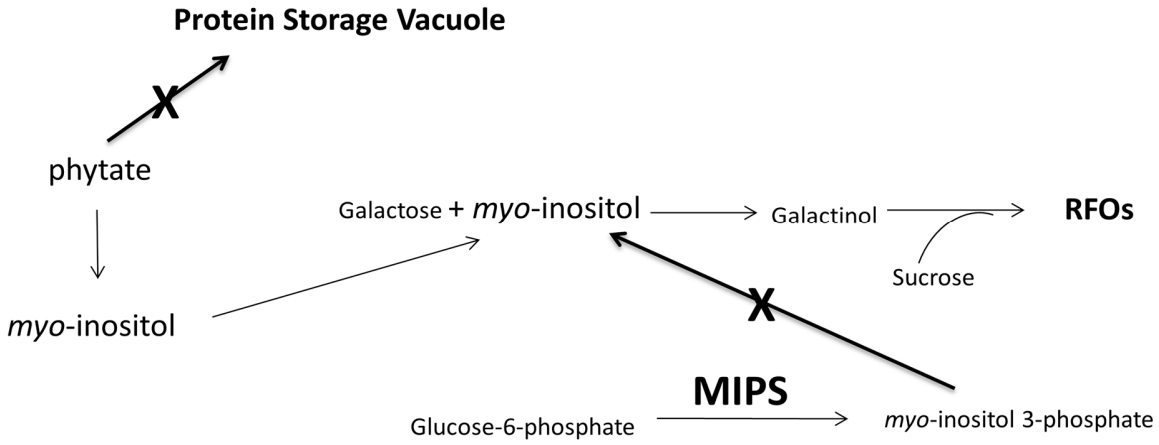


Figure 4.4. Proposed model for the interaction between the MIPS and MRP genes in V99-5089 x CX1834. MIPS is defective, therefore *myo*-inositol is not being produced, leading to low RFOs. In CX1834, phytate is unable to be transported into the protein storage vacuoles due to mutation in MRP genes. When looking at the three genes combined, this phytate gets broken down into its precursors, *myo*-inositol and lower-order intermediates and phosphates. This *myo*-inositol feeds into the RFO pathway, replacing the job of MIPS and resulting in high RFOs.

Table 4.1. Genotypic classes of the V99-5089 x CX1834 population, with their corresponding mean phytate (F_7 generation), sucrose, raffinose, and stachyose (F_8 generation) values, standard deviations, and ranges. Genotypic classes are designated in order of: MIPS1 allele/MRP-L allele/MRP-N allele. Lower case letters indicate mutant alleles and upper case letters indicate wild-type alleles.

PARENTAL MEANS													
		Phytate (mg g^{-1})			**Sucrose (%)**			**Raffinose (%)**			**Stachyose (%)**		
V99-5089		10.62			12.50			0.53			.25		
CX1834		8.78			7.14			0.73			4.48		
WHOLE POPULATION DESCRIPTIVE STATISTICS													
Genotypic class	**N**	**MeanÀ (mg g^{-1})**	**S.D.**	**Range**	**Mean (%)**	**S.D.**	**Range**	**Mean (%)**	**S.D.**	**Range**	**Mean (%)**	**S.D.**	**Range**
mips/MRP-L/MRP-N	15	10.74b	2.24	7.84-14.78	9.67a,b	2.21	5.89-12.78	0.49b	0.25	0.18-0.87	1.46b	1.73	0.13-4.31
mips/MRP-L/mrp-n	14	10.30b	1.88	8.15-14.00	11.16a	1.57	7.60-13.65	0.42b	0.14	0.25-0.68	0.54b	0.61	0.14-2.28
mips/mrp-l/MRP-N	20	9.95b	2.20	6.85-16.89	10.45a,b	1.69	6.6-13.51	0.41b	0.22	0.22-0.98	0.86b	1.30	0.19-4.27
mips/mrp-l/mrp-n	14	6.72c	2.31	4.01-11.96	8.87b,c	1.39	7.35-11.95	0.91a	0.34	0.43-1.71	3.64a	1.44	0.41-5.25
MIPS/MRP-L/MRP-N	29	14.60a	2.38	6.80-17.24	7.13d	1.02	6.02-10.59	0.83a	0.17	0.31-1.25	3.91a	0.88	0.17-5.10
MIPS/MRP-L/mrp-n	25	14.42a	1.29	11.53-17.33	7.36c,d	1.58	5.46-12.72	0.82a	0.15	0.51-1.07	3.87a	0.90	0.26-4.82
MIPS/mrp-l/MRP-N	29	14.13a	2.30	6.76-17.83	7.45c,d	1.36	5.78- 11.53	0.79a	0.18	0.26-1.07	3.46a	1.06	0.14-4.53
MIPS/mrp-l/mrp-n	17	6.94c	2.44	4.14-11.22	7.29c,d	1.07	4.52- 9.27	0.95a	0.12	0.82-1.29	4.25a	0.65	3.16-5.49

ÀMean phytate, sucrose, raffinose, and stachyose for all the lines in each genotypic class. For each trait, genotypic classes with values significantly different from one another are indicated by different letters (a-d), as determined by a Tukey-Kramer HSD comparison of means. Standard deviation (S.D.) for each genotypic class is listed next to the mean.

Table 4.2. Recombinant inbred lines in the mips/mrp-l/mrp-n genotypic class of the V99-5089 x CX1834 population. SNP and sequence data were collected from the F_8 generation of DNA. Oligosaccharide content data were collected by HPLC for the F_6, F_8, and F_{10} generations. Phytate data were collected for the F_7, F_8, F_9, and F_{10} generations. For the B1 MIPS1 sequence, the mutation is ñGò and the wild type is ñC.ò For the MRP L sequence, the mutation is ñA,ò and wild type is ñG.ò For the MRP N sequence, the mutation is ñTò and wild type is ñA.ò All fourteen lines were confirmed to each have all three mutations.

08 row #	B1 seq	L seq	N seq	F_6 Suc (%)	F_6 Raf (%)	F_6 Sta (%)	F_8 Suc (%)	F_8 Raf (%)	F_8 Sta (%)	F_{10} Suc (%)	F_{10} Raf (%)	F_{10} Sta (%)	Phy F_7	Phy F_8	Phy F_9	Phy F_{10}
819	G	A	T	8.09	0.90	3.76	8.06	0.77	4.20	11.28	1.14	6.67	5.74	5.09	3.96	4.97
821	G	A	T	8.09	0.62	3.22	5.89	0.55	2.87	7.85	1.40	6.71	6.45	ND	5.90	5.59
831	G	A	T	8.29	0.68	2.78	8.23	0.89	4.09	7.05	1.47	6.10	7.96	6.53	6.26	6.52
833	G	A	T	7.56	0.66	3.85	8.27	0.88	4.61	5.74	0.99	5.33	5.21	5.43	5.12	6.19
853	G	A	T	7.27	0.80	3.42	6.96	0.72	3.50	6.84	1.07	4.97	5.03	ND	5.35	4.43
859	G	A	T	6.86	0.68	3.04	6.43	0.71	3.18	6.35	1.03	5.48	6.82	ND	6.1	4.92
866	G	A	T	7.96	0.70	3.32	7.83	0.80	5.12	6.52	1.10	5.69	5.83	7.85	7.59	5.19
884	G	A	T	8.23	0.84	3.32	9.17	1.10	3.48	10.45	1.14	5.43	4.01	4.5	6.52	2.41
886	G	A	T	7.68	0.81	4.03	8.81	0.94	5.25	6.75	0.90	6.40	4.51	4.6	5.43	4.50
903	G	A	T	ND[À]	ND	ND	7.35	0.96	3.50	5.30	1.48	5.26	6.54	5.4	5.59	6.15
944	G	A	T	8.90	0.74	2.29	10.97	0.43	1.53	9.84	0.34	1.50	6.95	7.05	8.39	10.49
959	G	A	T	9.88	0.62	2.63	11.95	0.47	0.41	8.23	0.89	5.09	11.23	7.89	8.67	3.90
998	G	A	T	7.43	0.88	3.30	8.79	1.71	3.51	6.70	0.72	4.48	11.96	5.6	6.87	3.33
1006	G	A	T	7.98	0.98	3.64	8.16	1.10	4.16	ND	ND	ND	5.85	4.67	5.58	ND

ÀñNDò = no data.

Supplementary Table 4.1. Descriptions of all eight genotypic classes.

B1 allele	**L allele**	**N allele**	**Definition**
mips	MRP-L	MRP-N	B1 mutation, L wild type, N wild type
mips	MRP-L	mrp-n	B1 mutation, L wild type, N mutation
mips	mrp-l	MRP-N	B1 mutation, L mutation, N wild type
mips	mrp-l	mrp-n	B1 mutation, L mutation, N mutation
MIPS	mrp-l	mrp-n	B1 wild type, L mutation, N mutation
MIPS	MRP-L	mrp-n	B1 wild type, L wild type, N mutation
MIPS	mrp-l	MRP-N	B1 wild type, L mutation, N wild type
MIPS	MRP-L	MRP-N	B1 wild type, L wild type, N wild type

V. CONCLUSIONS

In this study, we shed some light on the genetic basis of low phytate in soybean lines CX1834 and V99-5089. First, we used molecular mapping and the whole genome sequence to identify two MRPs as candidate genes for causing low phytate in CX1834. We used the whole genome sequence to delineate the causative mutations in both MRP genes. In this study we also evaluated emergence and phytate in a population derived from CX1834 and found emergence to be dependent on several factors, including genotype and environment. Lastly, we looked at the effect of various combinations of mutant alleles from both V99-5089 and CX1834 and observed the effect on phytate and oligosaccharide levels.

CX1834 is a low phytate line which was previously shown to have two low phytate QTL on linkage groups (LGs) L and N. In this research, we identified the genetic basis of the low phytate trait in CX1834 by using the soybean genome to locate candidate genes in the low phytate QTL regions in CX1834. In order to do this, we first mapped the four soybean MIPS genes to find out if they mapped to the low phytate QTL regions on LGs L and N. Because they did not, we ruled them out as candidate genes. Using the maize *lpa1* low phytate gene, we identified two soybean MRP ABC transporters with sequence homology to the maize gene. These two MRP genes mapped to the low phytate QTL regions on LGs L and N. Sequencing the genes from CX1834 and comparing them to the Williams 82 reference whole genome sequence revealed a SNP in the MRP gene located on LG N (causing a stop codon), and a SNP in the MRP gene located on LG L (causing an amino acid change of an arginine to a lysine). To confirm that these two mutations were actually the cause of low phytate in CX1834, we then sequenced 16 other parental lines and found that none of them had mutations at either of the SNP locationsð these mutations are unique to CX1834. We looked at various low phytate and high phytate

selections from the CX1834 x V99-3337 population and confirmed that those individuals with low phytate had the two mutations coming from CX1834.

One major disadvantage with low phytate soybeans, specifically CX1834, is the low seedling emergence. Emergence is a complex trait, influenced by a number of factors including genotype and environment. In the next part of this research, we used the recombinant inbred population CX1834 x V99-3337 to evaluate emergence and phytate on a population-wide level (245 lines). In order to answer the questions: Is emergence influenced by environment, genotype, or both? And, does low phytate necessarily mean low emergence? The population was planted in two locations (Kentland Farm, VA and Mt. Holly, VA) in two years (2009 and 2010). We sent the population to be genotyped for two SNPsð the causative mutations in CX1834, one on LG L and one on LG N. We also collected emergence data using seeds from each year and location. Emergence data was collected using the extended cold test at Iowa State, and this assay was chosen because of its high correlation with field emergence. The phytate assay was done at Virginia Tech using the modified colorimetric assay. After obtaining descriptive statistics on each data set, we determined that 2009 Kentland Farm and 2010 Kentland Farm mean RIL phytate over the population as a whole were not significantly different from each other, but significantly lower than 2009 Mt. Holly and significantly higher than 2010 Mt. Holly phytate. For emergence, in 2009 the mean emergence was slightly higher in Kentland Farm than Mt. Holly, but was very poor in both locations in 2010, with some lines having 0% emergence. Over all years and locations, emergence was positively and significantly correlated to phytate.

A standard least squares model determined that variation among year, location, genotypic class, year x genotypic class, and year x location interactions were all significantly affecting

emergence. Using the SNP genotypic data, we could evaluate phytate and emergence based on the different combinations of the SNP mutations. The class with both L and N low phytate alleles coming from CX1834 had significantly lower phytate and emergence than the other genotypic classes. Those individuals having either or none of the CX1834 mutations had no significant difference in their phytate or emergence levels.

The other low phytate line used in this study, V99-5089, was developed at Virginia Tech. In addition to being low phytate, it has high sucrose and low raffinosaccharide content. Previous research from our lab showed V99-5089 to have a SNP mutation in a *myo*-inositol phosphate synthase (MIPS1) gene location on LG B1, which causes low phytate and raffinosaccharide content. The final objective of this study was to use the recombinant inbred population derived from CX1834 x V99-5089 to observe the combinations of all three mutations and to see how different combinations of the mutations impact phytate and raffinosaccharide levels. This CX1834 x V99-5089 population is an advanced RIL population with 173 lines. We phenotyped the population for phytate and oligosaccharide content, in addition to collecting SNP data for all three mutations (B1, L and N). Because lower stachyose and phytate content in a soybean line should provide better nutritional value and metabolizable energy in animal feed, we wanted to know if it was possible to get even lower levels of phytate and RFOs when combining all three mutations. We found that the individuals with all three mutations, as well as those with just the MRP mutations combined, had the lowest phytate compared to the other genotypic classes. When looking at stachyose, all classes with the MIPS mutant allele had a low stachyose mean, and all classes with the MIPS wild-type allele had a high stachyose meanð except for the class with all three mutations combined. Since this class had the B1 MIPS mutation, we expected that it would be low stachyose.

These puzzling results caused us to take a closer look at the class having all three mutations. Since we thought maybe there could be a mistake in genotyping or phenotyping, we sequenced all the lines to confirm that they indeed have all three mutations. We also collected oligosaccharide and phytate data for several subsequent generations. In all cases, the lines with all three mutations had low sucrose, and high stachyose. Since we still obtained these unexpected results, we tried to explain this by taking a closer look at the interaction among the three low phytate genes in this population: MIPS, and two MRPs. As stated earlier, in V99-5089, MIPS is defective, therefore *myo*-inositol is not being produced, leading to low RFOs. In CX1834, phytate is unable to be transported into the protein storage vacuoles due to mutations in MRP genes. Our hypothesis is: when looking at the three genes combined, this free phytate in the cytoplasm gets broken down into its precursors, *myo*-inositol and lower-order intermediates. This *myo*-inositol could feed into the RFO pathway, replacing the role of MIPS. This may explain the high stachyose levels observed with all three mutations even though MIPS is non-functional.

This research exemplifies the use of the soybean whole genome sequence (WGS). At the time of these studies, a draft of the soybean WGS had recently been released online. We used the WGS to identify candidate genes responsible for low phytate in CX1834 by doing a Blast homology search with the maize *lpa1* gene. We also used the WGS to design new primers for mapping MIPS and MRP genes. It was possible to search for SSRs very tightly linked to the genes of interest. This was especially useful when primers designed to amplify SSRs within the genes did not display any polymorphism. For this research, and for future research, the WGS can be used to design perfect markers to be used for genotyping important agronomical traits for marker-assisted selection.