

Evaluating the Utility and Causative Genetics of Soybean Mutants

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Benjamin W. Campbell

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Robert M. Stupar & James H. Orf

March, 2016

© Benjamin W Campbell 2016

Acknowledgements

I would like to express my gratitude to my co-advisors Drs. Robert Stupar and James Orf as you have been excellent mentors for me. Thank you for all of your encouragement, wisdom, and for providing the right balance of both guidance and freedom in my research. I would also like to thank my committee members, Drs. James Anderson, Seth Naeve, Carroll Vance, and Gary Muehlbauer for serving as my committee members. I would also like to acknowledge the assistance of the many faculty, staff, post-docs, and fellow graduate students who greatly assisted this research.

A special thank you to my family and friends. I am very grateful for my parents' support and encouragement throughout my graduate work. I would also like to thank all of my friends who supported me in my research. Words fail to express my appreciation for my beloved wife Megan as she has been a constant source of love, encouragement, comfort, and support.

I would also like to acknowledge the 'revealer of mysteries' for giving us knowledge of what He has created and without whom I could do nothing.

ABSTRACT

Soybean (*Glycine max* (L.) Merr.) is the second most widely planted crop in the United States by acreage, but yet its genetic resources, mapping methodologies, and breeding improvements lag behind those of other major crop species. In the 20th century, soybean researchers gathered a wealth of natural soybean genetic diversity in the forms of soybean's wild relative *G. soja*, soybean landraces, soybean elite lines, and spontaneous mutants. Starting in that same century, researchers began inducing soybean mutations through chemical or irradiation mutagenesis to generate new phenotypes. In the 21st century, these mutagenesis efforts have expanded and have been coupled with new genomics tools to enhance soybean functional genomics. These new mutagenesis efforts and genomics tools will be discussed in chapter one.

One of the challenges facing soybean is the difficulties in gene mapping, cloning, and validation. A major focus of this dissertation is the adaptation of new genomics tools and mapping methodologies to soybean in order to facilitate the identification of causative mutants in soybean. Chapter two demonstrates a more classical approach to gene mapping and soybean whole plant transformation to identify the causative loci for three spontaneous chlorophyll deficient mutants. In contrast, chapter three utilizes a combination of new genomics approaches to map and clone a fast neutron induced mutant and validates the result using both a second mutant allele from a historic soybean mutant and transformation of an *Arabidopsis* mutant. Chapter four builds off of the

results of chapter three in leveraging the genomic mapping approach to clone a spontaneous canopy architecture mutant.

Several unexpected results and conclusions are reported in the following chapters. Chapter two provides evidence to challenge the widely held idea of gene redundancy in soybean provides an effective buffer against mutations. Additionally, to our knowledge, the research of chapter two reports the first instance of identical mutations affecting two different paralogs resulting in nearly identical phenotypes. Chapter three demonstrates that array comparative genomic hybridization technology and whole genome sequencing of mutant and wild-type bulks can be effectively combined to map and clone a fast neutron mutant from a small F₂ population. The chapter also provides an example of the high complexity of mutations that can result from fast neutron irradiation. Chapter four describes the mapping and characterizing a short petiole mutant. The research identifies that the short petiole trait (*lps1*) is due to a three base-pair in frame insertion in an uncharacterized gene. It was found that the mutation decreases petiole length primarily by decreased cell length and that the short petiole trait could be agronomically beneficial through improved harvest index. The results from chapter four suggest that there is the capacity to improve soybean's productivity and agronomics through modifications to canopy architecture, as has been demonstrated in other major crop species.

The fifth and final chapter discusses potential future directions for soybean genomics research. New population designs with improved efficiency are described. Additionally, suggestions are made for how to utilize current technologies to improve next generation population designs.

Table of Contents

ABSTRACT.....	ii
List of Tables	viii
List of Figures	ix
CHAPTER 1	1
Soybean Mutant and Germplasm Resources: Current Status and Future Prospects	1
SUMMARY	1
INTRODUCTION	2
Sources of Natural Genetic Diversity	2
Sources of Induced Genetic Variation	7
COMMUNITY ACCESSIBLE SOYBEAN MUTATION RESOURCES	10
USDA Isoline Collection and USDA Soybean Genetic Type Collection.....	10
Chemical mutagenesis populations.....	11
Irradiation mutant populations	12
Transposon mutagenesis populations.....	14
<i>Ac/Ds transposon mutant population</i>	16
<i>mPing transposon mutant population</i>	18
<i>Tnt1 transposon mutant population</i>	20
<i>Tgm9 transposon mutant population</i>	22
BEST PRACTICES AND RECOMMENDATIONS FOR DEVELOPING SOYBEAN MUTANT POPULATIONS	23
1. Select a founder line and a seed lot.....	24
2. Choose a mutagen	26
3. Determine the size of the M ₁ and M ₂ populations	27
4. Select a M ₁ to M ₂ Advancement procedure	28
5. Screen and harvest M ₂ individuals.....	29
6. Advance future generations	29
7. Long term seed storage	30
CONCLUSIONS AND FUTURE PROSPECTS	31
CHAPTER 2	33
Identical Substitutions in Magnesium Chelatase Paralogs Result in Chlorophyll-Deficient Soybean Mutants.....	33
SUMMARY	33

INTRODUCTION	34
MATERIALS AND METHODS.....	36
Discovery and phenotype analysis of the MinnGold mutant	36
Genetic mapping of chlorophyll deficiency in MinnGold	36
Identification of Glyma13g30560 as a candidate for <i>y11</i>	38
Sequencing of Glyma13g30560 and Glyma15g08680 in mutant and wild-type lines	38
CAPS assays for <i>y11</i> and CD-5	40
Transformation and transgene analysis.....	41
Analysis of sequence similarity between the soybean ChII subunits.....	43
RESULTS	44
Identification, mapping, and fine-mapping of the MinnGold mutation.....	44
Identification of the candidate gene	45
Transgenic validation of the ChIIa Mg chelatase subunit.....	46
Sequence analysis suggests the mutation in MinnGold is allelic to <i>y11</i>	47
Identification of identical yellow foliage mutations at paralogous magnesium chelatase genes.....	48
Sequence comparison of the Mg-chelatase subunit across photosynthetic species	49
DISCUSSION.....	50
CHAPTER 3	56
Fast Neutron induced structural rearrangements at a soybean <i>NAPI</i> locus result in <i>gnarled</i> trichomes.....	56
SUMMARY	56
INTRODUCTION	57
MATERIALS AND METHODS.....	60
Populations and Phenotyping.....	60
Detection of structural variants using comparative genomic hybridization microarrays.....	60
Sequencing of R55C01, ‘Noir 1’, and F ₂ bulks	61
Whole Genome Sequencing based Bulk Segregant Analysis (WGS-BSA).....	62
Genotyping and Phenotyping of segregating F ₃ individuals.....	63
Validation of chromosome rearrangements in <i>GmNAPI</i>	63
RNA sequencing of R55C01 and ‘M92-220’ root, seed, and leaf tissue.....	64
Mapping the <i>p2</i> introgression interval	65
Sequencing the <i>GmNAPI</i> gene in T31.....	65
Transgenic complementation of Arabidopsis <i>glr-4</i> using <i>GmNAPI</i>	66

RESULTS	67
Identification and mapping of the <i>gnarled</i> trichome mutant	67
Complex rearrangements detected in Glyma.20G019300	69
RNA sequencing expression analysis of the <i>gnarled</i> mutant.....	71
Complementation of the <i>Atmap1</i> using <i>GmNAPI</i>	73
Identification of a spontaneous <i>NAPI</i> soybean mutant T31 (<i>p2</i>)	74
DISCUSSION.....	75
CHAPTER 4	79
Map-based cloning and physiological characterization of the soybean <i>lps1</i> short petiole locus ...	79
SUMMARY	79
INTRODUCTION	80
MATERIALS AND METHODS.....	82
Mapping population development	82
Whole Genome Sequencing based Bulk Segregant Analysis	83
Population advancement From F _{2:3} to F _{3:4}	84
Fine mapping of the <i>lps1</i> locus.	84
Identification of the candidate polymorphism	85
Segregation test of the <i>lps1</i> mutant with mutant phenotype	86
Selection and advancement of F _{3:4} individuals to development F _{4:6} NILs for preliminary yield trials	87
F _{4:6} Preliminary yield trials	87
Advancement of plant rows from F _{3:4} to F _{5:6}	88
F _{5:6} Near Isogenic Lines.....	89
RESULTS	90
Gross Mapping of <i>lps1</i> mutant to Chromosome 16	90
Fine Mapping of the <i>lps1</i> mutation	92
Candidate gene identification	92
Segregation test of the <i>lps1</i> mutant with mutant phenotype	93
Candidate gene characterization	94
Agronomic comparisons of <i>Lps1</i> and <i>lps1</i> NIL pairs	96
Yield and quality comparisons.....	97
Physiological characterization of <i>lps1</i> mutants.....	99
DISCUSSION.....	100

CHAPTER 5	107
Future Directions	107
ILLUSTRATIONS	110
FIGURES.....	110
Chapter 1 Figures.....	110
Chapter 2 Figures.....	110
Chapter 3 Figures.....	120
Chapter 4 Figures.....	128
Chapter 5 Figures.....	134
TABLES	135
Chapter 1 Tables	135
Chapter 2 Tables	138
Chapter 3 Tables	142
Chapter 4 Tables	144
Chapter 5 Tables	155
BIBLIOGRAPHY.....	156

List of Tables

Table 1. Soybean mutant populations created using chemical, irradiation, or transposon mutagenesis.....	135
Table 2. Basic summary information for different types of transposons used in soybean mutagenesis.....	137
Table 3. PCR Primers used to amplify Glyma13g30560 for <i>y11</i> and <i>y11-2</i>	138
Table 4. PCR Primers used to amplify Glyma15g08680 for CD-5.	139
Table 5. Twenty-two genes models present in the fine-mapped interval Gm13: 33,141,206..33,306,556.....	140
Table 6. Soybean CHLI Genes and Expression Data.	141
Table 7. Codominant PCR primer triple used to amplify mutant and wild-type alleles.	142
Table 8. PCR Primers used to amplification across chromosome rearrangements in Glyma.20G019300.....	142
Table 9. PCR Primers used to sequence Glyma.20G019300 in line T31 (PI548159)....	142
Table 10. PCR Primers used to test for the presence of the <i>GmNAP1</i> construct in Arabidopsis T ₁ individuals.....	143
Table 11. Primers and enzymes used to test candidate <i>lps1</i> polymorphisms.	144
Table 12. D76-1609 and its parental haplotypes in and adjacent to the <i>lps1</i> fine mapped interval.	145
Table 13. NIL average vegetative and reproductive growth stages and average plant height.....	146
Table 14. Preliminary yield trial results from <i>Lps1</i> and <i>lps1</i> Near Isogenic Line yield trials.....	147
Table 15. Additional data from preliminary yield trial results from <i>Lps1</i> and <i>lps1</i> Near Isogenic Line yield trials.....	151
Table 16. Examples of comparisons made between <i>Lps1</i> and <i>lps1</i> NIL pairs.	154

List of Figures

Figure 1. Phenotypic evaluation of chlorophyll deficiency in the MinnGold mutant. ...	110
Figure 2. Fine mapping the MinnGold locus.	111
Figure 3. The <i>y11-2</i> and <i>y11</i> mutations in the candidate gene Glyma13g30560 appear to be novel <i>de novo</i> mutations.	112
Figure 4. A Glyma13g30560 wild-type transgene complements the chlorophyll deficiency phenotype.	113
Figure 5. Phenotypic classes for chlorophyll deficiency mutants.	113
Figure 6. <i>y11</i> CAPS assay.	114
Figure 7. A Cleaved Amplified Polymorphic Sequences (CAPS) assay of seventeen individuals segregating for the presence of the candidate CD-5 SNP.	114
Figure 8. The CD-5 mutation in the candidate gene Glyma15g08680 appears to be a novel <i>de novo</i> mutation.	115
Figure 9. Polypeptide sequence alignment of the Mg-chelatase ChII subunit for the interval surrounding the <i>y11</i> , <i>y11-2</i> , and CD-5 mutations.	116
Figure 10. Amino acid sequence comparison of ChIIa (Glyma13g30560) to ChIIb (Glyma15g08680).	117
Figure 11. Speculated combinations of <i>y11-2</i> (A), <i>y11</i> (B), and CD-5 (C) mutant and wild-type ChIIa and ChIIb subunits arranged into hexameric rings.	118
Figure 12. The inferred evolutionary history for the four Glycine Max ChII subunits calculated using the UPGMA method.	119
Figure 13. The mutated Glyma.20G019300 allele co-segregates with the <i>gnarled</i> phenotype.	120
Figure 14. Phenotype and course genetic mapping of the <i>gnarled</i> trichome mutant.	121
Figure 15. Genetic mapping of the <i>gnarled</i> mutant and physical mapping of the deletion on chromosome 20.	122
Figure 16. Mutations in the candidate gene demonstrate the complexity of mutations that can occur by fast neutron mutagenesis.	123
Figure 17. RNA-seq read alignment density for each exon of Glyma.20G019300 in wild-type and <i>gnarled</i> mutant plants.	124
Figure 18. Soybean <i>GmNAP1</i> functionally complements Arabidopsis <i>nap1</i> mutant (<i>grl-4</i>).	125
Figure 19. <i>p2</i> introgression interval identified on chromosome 20.	126
Figure 20. Sequence comparison of T31, ‘Williams 82’ (Wm82), and 25 diverse wild-type lines for the interval flanking the frame shift deletion found in T31.	127
Figure 21. Short petiole phenotype of <i>lps1</i>	128
Figure 22. F _{2,3} <i>lps1</i> Bulk Segregant Analysis.	128
Figure 23. <i>lps1</i> Fine mapping interval and candidate gene.	129
Figure 24. Uniqueness of the candidate <i>lps1</i> mutation in soybean germplasm.	130
Figure 25. Petiole length is affected by environmental conditions.	130

Figure 26. Expression differences between the <i>lps1</i> candidate gene (Glyma.16g209100) and its paralog (Glyma.09g159900).	131
Figure 27. Preliminary yield trial results for short petiole versus long petiole Near Isogenic Lines.	132
Figure 28. Petiole lengths and cell lengths from paired Near Isogenic Lines.	133
Figure 29. Predicted effects of petiole length on leaf angle of attachment and light penetration into the soybean canopy.....	134

CHAPTER 1

Soybean Mutant and Germplasm Resources: Current Status and Future Prospects

SUMMARY

Genetic bottlenecks during domestication and modern breeding limited the genetic diversity of soybean (*Glycine max* (L.) Merr.). Therefore, expanding and diversifying soybean genetic resources is a major priority for the research community. These resources, consisting of natural and induced genetic variants, are valuable tools for improving soybean and furthering soybean biological knowledge. During the 20th century, researchers gathered a wealth of genetic variation in the forms of landraces, *Glycine soja* accessions, *Glycine* tertiary germplasm, and the USDA Type and Isoline Collections. During the 21st century, soybean researchers have added several new genetic and genomic resources. These include the reference genome sequence, genotype data for the USDA soybean germplasm collection, next-generation mapping populations, new irradiation and transposon-based mutagenesis populations, and designer nuclease platforms for genome engineering. This paper briefly surveys the publicly accessible soybean genetic resources currently available or in development and provides recommendations for developing such genetic resources in the future.

INTRODUCTION

Genetic variation is the foundation of all plant breeding programs, and thus the discovery, maintenance, and development of genetic diversity is essential to achieve sustained future yield gains to meet growing population demands in the face of climate change. There are many sources of genetic diversity that modern breeders and geneticists can utilize for soybean (*Glycine max* (L.) Merr.) improvement and for gaining insight into soybean biology. The introduction describes some of the natural and induced sources of genetic variation in soybean while the middle portion of the review focuses primarily on the development, application, and findings of soybean mutant populations. Additionally, we describe a list of recommended ‘best practices’ for generating and studying soybean mutant populations based on lessons learned from the past.

Sources of Natural Genetic Diversity

Soybean is an ancient polyploid species that experienced its most recent whole-genome duplication event approximately 13 million years ago (Schmutz et al., 2010). The genome has retained about 75% of its predicted genes as two or more copies; however, the extent of functional redundancy among paralogs and different gene families is not clear. This redundancy raises important considerations for scientists interested in developing and characterizing soybean genetic resources, influencing everything from mutagenesis strategies to phenotyping approaches. So while this paper does not focus on the effects of gene redundancy in mutant and germplasm resources, researchers are

encouraged to consider the impacts of gene redundancy when generating and evaluating these resources, as this redundancy can greatly influence the outcomes of specific projects and traits.

G. max began to take on a form that is clearly distinct from wild soybean (*Glycine soja* Sieb. and Zucc.) starting with an estimated single domestication event (Guo et al., 2010; Zhao et al., 2015; Zhou et al., 2015) that occurred approximately 3,000 years ago (Carter et al., 2004; Hymowitz and Shurtleff, 2005). Several prominent developmental features distinguish *G. max*. (yellow seeds, larger seeds, decreased shattering, and erect growth) and *G. soja* (black seeds, smaller seeds, shattering pods, and prostrate growth) (Liu et al., 2007; Li et al., 2013). These fundamental physiological changes are estimated to have been caused by a small number of major quantitative trait loci (QTL) and potentially a large number of minor QTL (Liu et al., 2007; Li et al., 2013). While *G. soja* lacks the desirable agronomic traits of *G. max*, *G. soja* is a source of extensive genetic variation. The domestication bottleneck of *G. max* was severe, and it is estimated that approximately half of the genetic diversity in *G. soja* was lost during domestication, as estimated by nucleotide diversity (π) or by pairwise divergence per nucleotide θ_π (Nei and Li, 1979; Tajima, 1983; Zhou et al., 2015; Zhao et al., 2015). Thus *G. soja* is a valuable source of genetic diversity for soybean researchers. However, compared to the wild relatives of other crop species, such as teosinte (*Zea mays* ssp. *parviglumis*) (Wright et al., 2005) and wild-barley (*Hordeum vulgare* ssp. *Spontaneum*) (Morrell et al., 2005), *G. soja* has a much lower level of sequence diversity (Hyten et al., 2006).

Despite having less genetic diversity than other wild relatives, the ease of crossing between *G. max* and *G. soja* has allowed researchers to access this source of genetic diversity. Several studies have discovered *G. soja* QTL that enhance agriculturally important traits, such as seed protein level (Nichols et al., 2006), resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe) (Wang et al., 2001; Winter et al., 2007), salt tolerance (Lee et al., 2009), and yield (Sebolt et al., 2000; Kabelka et al., 2006; Concibido et al., 2003; Li et al., 2008; Lee et al., 2009). These studies have often required the researchers to conduct several rounds of backcrossing the *G. soja* line to a cultivated soybean line, with or without molecular markers, in order to develop *G. max* lines suitable for field evaluations that containing the *G. soja* QTL (Carpenter and Fehr, 1986).

Soybean lacks a secondary gene pool, thus researchers must look to the tertiary gene pool to access more genetic diversity. Crosses between *G. max* and the tertiary gene pool, which includes *G. tomentella*, *G. argyrea*, and *G. latifolia*, requires both backcrossing and the laborious step of embryo rescue (Singh and Nelson, 2015). Despite these challenges, intriguing results have been reported wherein traits from the tertiary gene pool have been introduced into *G. max*. For example, loci from *G. tomentella* conferring resistance to soybean cyst nematode (Riggs et al., 1998) and loci from *G. tomentella* conferring resistance to soybean rust (*Phakopsora pachyrhizi* Syd.) (Singh and Nelson, 2015) have each been introduced into soybean.

While the wild relatives of soybean provide a rich source of genetic diversity, they are also less accessible to many soybean researchers due to crossing boundaries and

linkage drag. In contrast, soybean landraces, also known as plant introduction (PI) lines, are a more readily accessible source of genetic diversity. This source represents a diverse set of genetic material that is one-step removed from wild germplasm, but also exhibits reduced levels of genetic diversity as compared to wild relatives (Zhao et al., 2015; Zhou et al., 2015). Despite this loss in diversity, PI lines are an attractive genetic resource to many soybean breeders as these lines are more easily crossed to elite material and progeny can be evaluated in the field with little to no backcrossing.

The historic value of PI lines to soybean breeders can be seen in North American soybean breeding pedigrees. Gizlice et al. (1994) estimated that from 1947 to 1988, 80 PIs provided the main source (99%) of genetic diversity for soybean breeding in North America, and the authors estimated that 75% of the genes in modern North American cultivars could be traced back to just 17 first progeny lines released before 1960 that were derived from crosses between PI lines. Still, the value of using PI lines to improve yield in modern elite lines is in question. Some researchers have found that across several breeding populations there is an inverse correlation between the population mean yield and the percentage of PI germplasm in the population, suggesting that PI lines are not a viable method of improving soybean yield in the short term (Schoener and Fehr, 1979; Vello et al., 1984; Ininda et al., 1996). In contrast, Thompson and Nelson (1998) found that they were able to produce progeny yielding greater than elite germplasm in both PI-by-PI crosses and PI-by-elite line crosses when using the PI lines that were genetically distinct from each other and from elite North American germplasm. The same study also found that higher yielding elite lines could be developed by backcrossing PI lines to elite

materials when the PI lines used were genetically distinct from the North American elite lines. These results suggest that breeders should carefully evaluate the genetic relationships between the PI lines and the current elite lines before selecting PI parents in specific breeding schemes. The recent genotyping of the USDA soybean germplasm collection provides breeders with high quality genotype data on thousands of PI lines (Song et al., 2015). This information can be used to guide the selection of PI parents in future crosses.

PI lines have also been used as an important source of resistance genes. For example, PI 88788 provided the valuable *rhg1-b* soybean cyst nematode (*Heterodera glycines*) resistance allele (Ross and Brim, 1957; Concibido et al., 1997; Cook et al., 2012) that is now used by breeders around the world. A short list of other valuable traits derived from landraces include increased oleic acid (Pham et al., 2010) and reduced linolenic acid (Mounts et al., 1988) in the seed, resistance to soybean rust (Hyten et al., 2009), and the absence of the Kunitz trypsin inhibitor (Hymowitz et al., 1978; Orf and Hymowitz, 1979).

In addition to wild-relatives and PI lines, soybean researchers have identified and maintained stocks of various classical soybean mutants in the Soybean Genetic Type Collection. The mutants in the Soybean Genetic Type collection are organized by an assigned designation ‘T’ followed by a number (Bernard, 1976). The vast majority of the mutants were naturally or spontaneously derived (Carter et al., 2004). Some of the mutants in the Soybean Genetic Type Collection have been backcrossed into primarily three cultivars (‘Clark’, ‘Harosoy’, and ‘Williams’) to generate a near isogenic line

collection called the Soybean Isoline Collection (Bernard et al., 1991). The Genetic Type Collection and the Isoline Collection continue to be valuable genetic resources for understanding soybean biology (Ainsworth et al., 2004; Cober et al., 2010; Tardivel et al., 2014; Campbell et al., 2014; Schmidt et al., 2015).

Next-generation mapping populations are also being developed within the soybean community. The recently completed soybean nested association mapping population (SoyNAM) (details available at Soybase (Grant et al., 2010); <http://www.soybase.org/SoyNAM/>) provides a valuable source of genetic diversity for researchers (Stupar and Specht, 2013). The genetic diversity present in the parental lines should provide a wealth of valuable data for future studies. One of the first papers looking at the SoyNAM parent diversity described the gene copy number and presence-absence differences among the parental lines (Anderson et al., 2014). Additionally, SoyNAM provides researchers with a developed set of genotyped recombinant inbred lines, which will allow future researchers to advance straight into phenotyping and mapping without having to first develop and genotype a segregating population. The collaborative efforts and sharing of data from multiple phenotypes further enhances the value of the SoyNAM resource.

Sources of Induced Genetic Variation

In addition to working with naturally occurring genetic variation, soybean researchers have purposefully generated genetic diversity in three ways: transgene insertion, targeted mutagenesis, and random mutagenesis. In 1996, the first

commercialized transgenic soybean product came onto the market and provided resistance to the herbicide glyphosate (Padgett et al., 1995). Additional transgenic events have been commercialized since then, including high oleic and low linoleic oil (Kinney, 1996; Mroczka et al., 2010; Brink et al., 2014), second generation glyphosate resistance (Horak et al., 2015), and dicamba herbicide resistance (Behrens et al., 2007). In addition, soybean geneticists have also used transgenes as tools to understand gene function or to demonstrate the potential value of certain transgenic applications (Falco et al., 1995; Stewart et al., 1996; Liu et al., 2012; Curtin et al., 2011; Cook et al., 2012; Campbell et al., 2014).

Site Specific Nucleases (SSN) are an exciting new suite of biotechnological tools that researchers are using to increase soybean genetic diversity through targeted DNA modifications. This field of research, known as genome engineering, uses SSN to generate double-stranded breaks (DSBs) at specific DNA sequences (target sites). The repair of these DSB generates new mutations or other DNA changes (e.g. specific nucleotide edits) at the target site. The most common SSN platforms include Meganucleases, Zinc Finger Nucleases (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs). SSNs have enabled soybean researchers to mutate specific genes to eliminate gene function, introduce specific nucleotide changes, and insert DNA sequences into specific locations in the genome (Michno et al., 2015; Haun et al., 2014; Jacobs et al., 2015; Li et al., 2015; Curtin et al., 2011). SSNs have proven particularly useful for mutating gene duplicates, allowing for the assessment of redundancies in

soybean gene function. For example, a desired high oleic acid seed phenotype was only recovered by mutating two specific fatty acid desaturase 2 gene paralogs (Haun et al., 2014). Furthermore, Curtin et al. (2015) recently demonstrated functional redundancy of two dicer-like 1 paralogs by SSN-based mutagenesis. Chen and Gao (2014) and Voytas and Gao (2014) have written more detailed reviews of plant SSNs and their applications. While still in their infancy, SSNs have demonstrated the powerful ability to introduce nearly unlimited genetic diversity into the soybean genome.

While transgenic and SSNs approaches have become popular for single gene analysis, efforts devoted to developing community-wide genetic resources for increasing soybean genetic diversity and for understanding soybean gene function have relied upon traditional mutagenesis. Random mutagenesis platforms, such as chemical and irradiation treatments, have been used by soybean researchers for several decades, and numerous studies have been published describing the development and utilization of soybean mutants to increase our understanding of soybean biology and to improve soybean quality. Furthermore, in recent years, efforts have been made to genetically transform plant transposons from other species into soybean, or to identify endogenous mobile elements in soybean, in order to generate new mutant populations. A survey of these diverse approaches and resources will be the primary topic of this article. While a great deal of interest surrounds the many different sources of soybean genetic diversity as described above, we will focus specifically on the publicly-available resources that have been developed or are in development from random mutagenesis projects and provide general recommendations for generating mutant population resources in the future.

COMMUNITY ACCESSIBLE SOYBEAN MUTATION RESOURCES

USDA Isoline Collection and USDA Soybean Genetic Type Collection

The USDA's Soybean Genetic Type Collection contains a wealth of soybean germplasm displaying unique qualitative genetic traits. As was mentioned in the "Sources of Natural Genetic Diversity" section, the strain nomenclature consists of a 'T' followed by a number (e.g. T31) that were assigned in chronological order of submission. The collection contains over 180 qualitative mutant strains for a variety of different traits that arose primarily by natural mutation, but a few mutants in the collection were generated by chemical or irradiation mutagenesis (each of these strains is summarized in Carter et al., 2004). The collection includes both dominant and recessive mutants, as well as recessive mutants that must be maintained as heterozygotes due to homozygous lethality or sterility.

Several of the USDA Soybean Genetic Type Collection strains were used to generate the Soybean Isoline Collection. The Soybean Isoline Collection is a diverse panel of almost 600 near-isogenic lines (NILs), developed by backcrossing mostly qualitative traits into primarily three cultivars: 'Clark', 'Harosoy', and 'Williams' (Bernard et al., 1991). Over 60 different nuclear genes and cytoplasmic factors were backcrossed into these parent lines, and many of the NILs containing the trait(s) were selected after five rounds of backcrossing (BC₅). With the exception of lethal or sterile traits, the NILs are maintained as homozygous lines. The diverse set of traits

encompassed in the Isoline Collection is organized under the following categories: Disease Resistance, Nutrient Response, Stem Growth, Time to Maturity, Combinations of Stem and Maturity Genes, Leaf Form, Pubescence Type, Chlorophyll, Pigmentation, Other, Combinations Transferred Together, and Miscellaneous Recombinations.

Chemical mutagenesis populations

There have been numerous studies that have utilized chemical mutagenesis to generate new genetic variation in soybean, and some of these studies are listed in Table 1. The two most commonly used chemical mutagens are Ethyl methanesulfonate (EMS) and N-Nitroso-N-methylurea (NMU). Cooper et al. (2008) demonstrated the proof of concept for conducting chemical mutagenesis and identification of mutations in soybean using a strategy called Targeting Induced Local Lesions IN Genomes (TILLING) (McCallum et al., 2000). These mutant populations have resulted in the identification of genes involved in disease resistance (*Rhg4*; Liu et al., 2012), seed composition traits (e.g. Dierking and Bilyeu, 2009; Gillman et al., 2013; Carrero-Colón et al., 2014; Hoshino et al., 2014), maturity and flowering time loci (Watanabe et al., 2009, 2011; Xia et al., 2012) as well as providing a wealth of knowledge about various biochemical and physiological processes (Table 1). While there is substantial community interest in further developing chemically mutagenized populations for soybean research and breeding (e.g. Tsuda et al., 2015), to our knowledge no such populations are currently considered “public” resources.

Irradiation mutant populations

Numerous studies have been conducted using irradiation of soybeans to generate mutants for forward genetics research as well as to generate new genetic variation for breeding (Table 1). Two publicly available soybean mutant populations were recently developed using fast neutron irradiation. One population was developed at the University of Minnesota using the line ‘M92-220’, which was derived from the cultivar ‘MN1302’ (Orf and Denny, 2004). The other population was developed at the University of Missouri using the cultivar ‘Williams 82’ (Bernard and Cremeens, 1988). Fast Neutron mutagenesis was once thought to primarily generate relatively small deletions (Li et al., 2001), but results from these populations identified both small and larger deletions, duplications, translocations, and inversions (Bolon et al., 2011, 2014; Findley et al., 2011; Gillman et al., 2014). Array Comparative Genomic Hybridization (CGH) has been used to detect large chromosomal deletions and duplications in these populations and in a variety of plant species (Bart et al., 2010; Ríos et al., 2008; Gong et al., 2004; Haun et al., 2011; Bolon et al., 2011, 2014; Gillman et al., 2014). Furthermore, there is evidence that fast neutrons can also generate smaller DNA sequence changes, including single base substitutions (Belfield et al., 2012).

To develop the ‘M92-220’ fast neutron population, seed batches were treated with one of four levels of fast neutron radiation (4, 8, 16, and 32 Gy) at the University of California-Davis’s McClellan Nuclear Radiation Center (Bolon et al., 2011). The irradiated seed was planted in the field and advanced by single seed descent. At the M₂

generation, more than 20,000 individual plants were visually phenotyped, tagged, tissue sampled, and M₃ seeds from each M₂ plant were collected for long term cold storage (Bolon et al., 2011). Currently 258 soybean fast neutron mutant lines have been genotyped by a CGH platform that can reliably detect deletions and duplications greater than 2kb in length (Haun et al., 2011). More than 5,000 fast neutron mutant individuals have been screened for root phenotypes using visual assessments, and more than 17,000 individual lines have been screened for seed composition phenotypes using a Perten DA7200 NIR Analyzer (Hägersten, Sweden). In addition to seed composition and root architecture mutants, several other mutant classes were identified: shoot, morphological, seed size, maturity, and feeding preferences of Japanese Beetles (*Popillia japonica*). The likely causative mutations for three fast neutron mutants have been identified from this population (Bolon et al., 2011, 2014), and the causative loci have been tentatively identified for several additional mutants that affect seed composition, trichome morphology, and canopy architecture. A core collection of just over 500 fast neutron mutant lines displaying the greatest genetic variation has been established to facilitate the ease of population maintenance and to concentrate subsequent genotyping and phenotyping efforts. Some of the core collection lines have CGH and/or resequencing data available. The phenotypic data and CGH genotype data are publicly available at Soybase (Grant et al., 2010), at the address <http://soybase.org/mutants/>.

The 'Williams 82' fast neutron mutant population was also developed by irradiating seed batches at the University of California-Davis's McClellan Nuclear Radiation Center. CGH has been performed on a subset of this population to identify

putative deletions and duplications in the mutant genomes. A specific deletion has been identified that causes both high stearic acid in the seeds and morphological alterations to root nodules (Gillman et al., 2014). Additionally, a mutant was identified that accumulates homogentisate which results in a brown seed coat phenotype. The increased homogentisate accumulation in the mutant also results in increased production of vitamin E and tolerance to herbicides that target homogentisate biosynthetic pathways (M.G. Stacey, personal communication). Lastly, a mutant line harboring a deletion for *Inositol Pentakisphosphate 2-Kinase (IPK1, Glyma06g03310)*, encoding an enzyme involved in phytic acid production, was also isolated. This novel mutation, on its own, produced wild-type levels of seed phytic acid. However, when this novel mutation was combined with the mutated *IPK1* isoform encoded on chromosome 14 (*Glyma14g07880*), a drastic reduction in seed phytic acid accumulation was observed. Unlike other low phytic acid (*lpa*) mutants in soybean, the double *IPK1* mutant showed no significant reductions in germination or field emergence (Vincent et al., 2015).

Transposon mutagenesis populations

In addition to chemical and irradiation mutagenesis populations, several soybean research groups have developed or are in the process of developing transposon mutagenesis populations. These projects are creating mutant populations using either endogenous transposons or transposons introduced from other plant species by genetic transformation.

Transposon mutagenesis populations provide both unique advantages and disadvantages as compared to chemical and irradiation mutagenesis populations. Unlike the large deletions caused by irradiation mutagenesis which may disrupt multiple genes, transposon insertional mutagenesis typically disrupts the function of single genes. Furthermore, flanking sequence data adjacent to the transposon can be used to identify the genomic locations of the inserted transposons and thus identify perturbed genes. The completion of the soybean genome sequence (Schmutz et al., 2010) and the availability of low cost sequencing allows for the localization of the transposons using methods such as sequence-capture, whole-genome sequencing, or thermal asymmetric interlaced (TAIL)-PCR (Liu et al., 1995). In addition to the aforementioned benefits, transgene-based transposon mutagenesis provides the opportunity to utilize more elaborate mutation tools such as enhancer traps and gene traps. Enhancer traps utilize a minimal constitutive promoter driving a visual marker gene that is only expressed when inserted in close proximity to an endogenous enhancer sequence (Martienssen, 1998). Gene trap designs also use a visual marker gene, but in this case, the inserted sequence contains a triple splice site acceptor and an intron upstream of the visual marker gene such that the visual marker gene is expressed if the gene trap element is inserted into the intron or exon of a gene (Martienssen, 1998). Thus, in addition to being able to perturb gene function by insertion, valuable data can be generated from both enhancer traps, through producing visual markers of gene expression patterns, and from gene traps, by causing mis-expression of the gene.

One interesting feature of the transposon element system is the possibility that some elements could continue to proliferate in both somatic and germline tissues. This proliferation can be viewed as an advantage in that additional whole plant transformation is not needed to increase the number of insertion events. However, some researchers may prefer a more stable mutation system that does not undergo *de novo* transposon mutagenesis during the phenotypic evaluation. The systems currently being developed in soybean exhibit a range of transposon activity following transformation, from very active to completely inactive.

Three transposon-based soybean populations currently being developed introduce well-characterized transposable elements from other plant species into soybean using genetic transformation. These include the *Ac/Ds*, *mPing*, and *Tnt1* transposon systems. One difficulty in utilizing this approach is that the resulting lines are transgenic and fall under the regulation of USDA-APHIS. Thus, the resulting mutant lines require additional operational considerations for field screenings and additional regulations for seed transfer across state lines. To avoid these obstacles, a soybean population based on transposition of the endogenous *Tmg9* element is also being developed. Plants derived from this population do not fall under USDA-APHIS transgenic regulation. Further details on these four soybean transposon populations are discussed below (Table 2).

Ac/Ds transposon mutant population

The *Activator/Dissociation (Ac/Ds)* transposon system (Jones et al., 1990; Izawa et al., 1997; Gidoni et al., 2003; Kolesnik et al., 2004) from maize was modified to

contain either enhancer or gene trap elements and subsequently an activation tag element was designed in which the cassava vein mosaic virus promoter is delineated by *Ds* termini. The *Ds* trap elements will mirror promoter activity, while the activation tags are useful for inducing mis-expression of tagged alleles (Mathieu et al., 2009). The *Ac/Ds* system allows the researcher to either stabilize a transposon insertion by removing the *Ac* element from a line, or to generate new insertions through the activation of the *Ds* element by maintaining the *Ac* element in a *Ds* containing lineage. Initially, the population design called for the *Ac* and *Ds* elements to be transformed into separate soybean lines and then combined together by crossing. By transforming the *Ac* and *Ds* elements separately into unlinked genomic locations, the researchers could easily activate or freeze the insertions.

Several enhancer trap lines were identified with GUS expression caused by transposon insertion into genes. Mathieu et al. (2009) demonstrated proof of concept for gene cloning using the soybean *Ac/Ds* transposon mutagenesis system by identifying that the cause of a male sterile mutant was the insertion of a T-DNA into a strictosidine synthase (STR) gene. The *Ac/Ds* population was found to have between one and five insertion events per line, with approximately 50% of the lines having a single insertion locus and 20% of the lines having two insertion loci. 200 insertion events were identified using flanking sequence data, and approximately 70% of the insertions disrupted known genes (Mathieu et al., 2009). As of October, 2015, the population had approximately 144 F₁ plants combining the *Ac* element with the *Ds*-enhancer trap element and approximately 550 F₁ plants combining the *Ac* element and the *Ds*-activation tag element. The current

emphasis is the creation of an activation tagged soybean population program targeting a repository of 3,000 to 4,000 mapped activation tags (Clemente, personal communication).

mPing transposon mutant population

A soybean mutant population using the *mPing* transposable element isolated from rice (*Oryza sativa*) (Jiang et al., 2003; Kikuchi et al., 2003; Nakazaki et al., 2003) is currently being developed (Hancock et al., 2011). *mPing* is a 430-bp element with terminal inverted repeats that has a high rate of transposition in rice (Naito et al., 2006). The non-autonomous *mPing* element is a derivative of the *Ping* element, as *mPing* lacks two open reading frames required for transposition (*ORF1* and *Transposase [TPase]*). Yang et al. (2007) found in Arabidopsis that the ORF1 and TPase proteins from either *Ping*, or the closely related *Pong* element, could cause the transposition of *mPing*, demonstrating an activator-dissociation transposon system similar to *Ac/Ds*. Hancock et al. (2011) repeated this finding in soybean by successfully freezing an *mPing* insertion through the removal of the *Ping* proteins.

By sampling multiple leaves from multiple plants in the T₁ and T₂ generations, Hancock et al. (2011) was able to demonstrate that the *mPing* event actively transposed throughout the soybean plant's growth and on average one new germinal *mPing* insertion was generated per generation. *mPing*'s active transposition throughout the plant's growth could generate additional somatic and germinal mutations that could complicate phenotypic analysis if these elements are not frozen by removing the *Ping* element. In cases when *mPing* does excise, Hancock et al. (2010) and Yang et al. (2007) found that

the excision sites were correctly repaired at frequencies of 99% in yeast (*Saccharomyces cerevisiae*) and 82% in Arabidopsis. In contrast, Hancock et al. (2011) found that in soybean approximately 70% of the excision events were repaired correctly; however, the remaining ~ 30% (7/23) of *mPing* excision events resulted in small deletions (7-23bp) (i.e. “footprints”). If this excision footprint alters gene function, then a mutant phenotype could be produced in both the line with an *mPing* insertion and in its sister line with only the excision footprint left from a subsequent transposition. Thus, the findings by Hancock et al. (2011) suggest that it is necessary to use sequencing to check for the presence of an excision footprint mutation when comparing the phenotypes of two lines for a particular *mPing* insertion: one line with an *mPing* insertion and its sister line lacking that same *mPing* insertion due to its subsequent transposition.

Two unique properties of *mPing* in soybean are its preferential insertion within 5 kb of a predicted gene transcript, and its ‘lack of avoidance’ of exon sequence (Hancock et al., 2011). Taken together, Hancock et al. (2011) demonstrated that the insertion characteristics of the *mPing* element make it ideal for insertional mutagenesis and speculated that *mPing* could also be ideal for activation tagging. Soybean populations containing versions of *mPing* designed for activation tagging and target silencing are being characterized. Currently the *mPing* mutagenesis population has been developed into approximately 1,400 lines grown over multiple generations. Mutant phenotypes have been identified in this population and hundreds of *mPing* insertion events have been mapped (Hancock and Kanizay, personal communication).

Tnt1 transposon mutant population

The *Tnt1* soybean mutant population uses the *Tnt1* retrotransposon element isolated from tobacco (*Nicotiana tabacum*) (Grandbastien et al., 1989; Cui et al., 2013). The *Tnt1* element has been effectively used in several plant species to generate mutants: *Arabidopsis* (Courtial et al., 2001; Lucas et al., 1995), lettuce (*Lactuca sativa*) (Mazier et al., 2007), and *Medicago truncatula* (Iantcheva et al., 2009; d'Erfurth et al., 2003; Tadege et al., 2005, 2008). These studies demonstrated that the *Tnt1* element is excellent for generating gene disruptions as it preferentially transposes into gene rich regions. Unlike the aforementioned *mPing* element, the *Tnt1* element does not transpose in normal plant growth conditions, but instead, the *Tnt1* element can be activated by passing *Tnt1* plants through tissue culture (Hirochika, 1993).

The activation of *Tnt1* in tissue culture could be problematic if a researcher is attempting to conduct a transgenic complementation test using a line carrying *Tnt1* insertions. If a researcher directly transforms a *Tnt1* plant, the *Tnt1* elements would likely generate new insertion mutations by retrotransposition, which could confound the complementation test. To avoid this problem, a researcher could first transform a line not carrying *Tnt1*, then cross the transgene of interest into the line containing the *Tnt1* insertion, and subsequently evaluate the phenotypes of the F₁ (heterozygous for both the *Tnt1* insertion and the transgene) and F₂ progeny (segregating for both the *Tnt1* insertion and the transgene). For this cross, an F₂ segregation ratio of 15:1 would be expected if the transgene complements the *Tnt1* insertion mutation and is inherited in a Mendelian fashion.

Cui et al. (2013) found that their *Tnt1* elements generate between four and 20 new insertions per plant each time it was activated in the soybean cultivar ‘Maverick’. Cui et al. (2013) used TAIL-PCR to identify the locations of 99 *Tnt1* insertions from 18 independent transgenic events. The authors found insertions on all 20 chromosomes and 62 of the 99 insertions (62%) occurred in annotated genes, which is higher than the random insertion rate of 9.8% calculated by the authors. These results indicate that *Tnt1* preferentially transposes into gene-rich regions of soybean and that *Tnt1* can transpose to all 20 chromosomes.

To identify if the *Tnt1* events continued to transpose or were stable outside of tissue culture, the authors used a *Tnt1* specific probe to conduct a Southern blot analysis to identify and compare the *Tnt1* locations in several T₀ lines and their resulting progeny. Additionally, for four *Tnt1* loci, the authors used PCR to generate amplicons that spanned from the *Tnt1* element into the flanking genomic sequence. From these tests, the authors found that the *Tnt1* events were stable and heritable. The PCR assay demonstrated that three of the four *Tnt1* insertions inherited in a Mendelian 1:2:1 fashion (homozygous wild-type: heterozygous: homozygous *Tnt1* insertion). The authors speculated that the *Tnt1* insertion that did not inherit in a Mendelian fashion could be due to lethality of the homozygous *Tnt1* insertion, as no individuals homozygous for that *Tnt1* insertion were recovered (0/18). Cui et al. (2013) demonstrated that they could generate new *Tnt1* insertions by reactivating the *Tnt1* element in tissue culture using both the method described by Zeng et al. (2004) using cotyledons from *Tnt1* containing T₁ plants as

explants or using the somatic embryo regeneration method described by Trick et al. (1997) using immature embryos collected from *Tnt1* containing T₁ plants.

Tgm9 transposon mutant population

While the introduction of *Ac/Ds*, *mPing*, and *Tnt1* elements into soybean each require genetic transformation, a recent population has been initiated to develop transposon mutants with an endogenous transposon from soybean, the *Tgm9* CACTA-type transposable element (Xu et al., 2010; Bhattacharyya, personal communication). As the *Tgm9* transposon is endogenous, the mutant population is not transgenic. Thus, while the *Tgm9* transposon population does not have gene trap or enhancer trap capabilities, the non-transgenic nature of this population provides researchers the benefit of working with a transposon mutagenesis system while avoiding transgenic regulatory complications.

The active *Tgm9* transposon was found in a line later designated *w4*-mutable (*w4-m*) due to its variegated flower color allele. Additional information about this line's origin was described by Groose et al. (1988) and Weigelt et al. (1990). Palmer et al. (1989) determined that the cause of the variegated flower color allele *w4-m* was due to transposon excision from the *W4* flower color locus. The authors also noticed that approximately 1% of the progeny from the plants with an active transposon also had mutations at other loci. Later, Xu et al. (2010) cloned the 20,548bp CACTA-like *Tgm9* transposon that had inserted into the *DRF2* gene creating the *w4-m* allele. Previous results have demonstrated the high transposition frequency of *Tgm9* in both germinal and

somatic tissues (Groose et al., 1988, 1990; Palmer et al., 1989). Palmer et al. (2008a, b) screened a collection of 3,206 *w4-m* plants looking for germinal revertants, identified by the presence of purple flowers (*W4*), indicating that the transposon had excised from the *w4-m* allele and inserted somewhere else in the genome. This screen yielded 36 independent male-sterile, female-fertile (MSFS) mutants and 24 independent root necrotic mutants, demonstrating the ability of this transposon to mutate other genes. Later, Raval et al. (2013) utilized the *Tgm9* transposon tagging system to identify an insertion within a DNA/RNA helicase/DEAD-box protein gene that caused a male-sterile, female-sterile mutant. Recently, it has been demonstrated that excision of the element from the gene resulted in restoration of the male and female fertility wild-type phenotype (Bhattacharyya, personnel communication).

Currently 5,184 lines have been sequenced from the *Tgm9* insertion sites, and there are over 17,000 lines with *Tgm9* insertions that are yet to be located (Bhattacharyya, personal communication). Preliminary results from that study and an earlier study of sequences from approximately 200 *Tgm9* insertion sites revealed that the element transposes to all 20 chromosomes and preferably to genic regions.

BEST PRACTICES AND RECOMMENDATIONS FOR DEVELOPING SOYBEAN MUTANT POPULATIONS

With all of the aforementioned genetic resources becoming available, the future prospects for soybean genomic research are bright. One can further envision a wider

array of genetic resources that will be developed in the future that are adapted to specific environments or for specific research applications.

The development and management of a mutant population requires considerable effort and careful decision-making throughout the process. Populations created for different purposes likely require vastly different development methods. For example, researchers developing a mutant population to be screened for resistance to a strong selection agent (e.g. an herbicide) might advance the population by a bulk harvesting method to M_2 and will maintain only a few surviving individuals after imposing the selection agent. In contrast, populations developed to serve as a community genetic resource will likely utilize a single seed descent (SSD) approach for several years and require the maintenance of thousands of lines. Here we will describe a list of general best practices and considerations for researchers when developing a long term, large mutant population for the purpose of forward genetic and/or reverse genetic screens.

1. Select a founder line and a seed lot

There are multiple items to consider when selecting a ‘founder’ line for mutagenesis, and the following items are not listed in any specific order. Are there certain genes (e.g. resistance genes) that are critical to mutate for a specific study? What is the founder line’s maturity group? Can the founder line be genetically transformed? How much pure seed is available for the founder line? Is the founder seed source completely homogeneous among individuals (as well as completely homozygous within individuals)? Is it important to have easily identifiable phenotypic markers to assist in

population purity and F_1 identification for future crossing applications? Does the founder line have good agronomic characteristics?

The following recommendations with respect to founder line selection may facilitate the identification and long-term utility of a given population:

- 1) A founder line that matures slightly earlier than is typical for your intended maturity zone will allow the harvesting of mutants with later maturities.
- 2) Utilizing a cultivar with grey pubescence facilitates rouging out any tawny pubescence individual contaminants or tawny outcrosses post-senesce and before harvest. For downstream mapping population development, it is helpful to have a combination of dominant and recessive phenotypic markers to visually identify successfully generated F_1 plants. Thus a researcher may wish to use a cultivar with grey pubescence and purple flowers to be able to outcross to a cultivar with either tawny pubescence or white flowers and still have the benefit of utilizing at least one dominant phenotypic marker.
- 3) The mutagenized seed stock should be directly descended from a single, inbred plant (at least F_8 , but preferably later to maximum homozygosity), possibly followed by two generations of seed increase (i.e. the mutagenized stock may be $F_{8:10}$ seed). In a post-genomics era, utilizing a seed lot that is both homozygous and homogeneous will facilitate the computational identification of induced genetic variation.
- 4) Maintaining seed purity is essential while increasing the stock of founder line seed. Actions that help maintain seed purity during the seed increase process

include: (a) using a field previously rotated out of soybean, (b) using clean planter equipment, (c) walking and rouging contaminants from the field during the season and before harvest, (d) threshing the seed using clean threshing equipment, and (e) storing the increased seed in clean and clearly labeled containers.

- 5) A founder line amenable to genetic transformation will be useful for gene validation experiments that use transgenic or genome engineering approaches.
- 6) A founder line with superior agronomic characteristics (e.g. good standability, disease resistance, high yield etc.) will facilitate the development and maintenance of the mutant population in the field. Additionally, a founder line with superior agronomics will make it easier for breeders to use derived mutant progeny for cultivar development.

2. Choose a mutagen

As previously mentioned, different mutagens produce different types of mutations, ranging from single base substitutions to large chromosomal rearrangements. Each type of mutagen will have advantages and disadvantages, and it is important to note that the type of mutagen used will affect downstream genotyping, mutation screens, and applications. For example, mutagenesis using transposons from different species may be excellent for exploring gene function, but transposon insertion mutants from this

population will be difficult to use for mutation breeding applications due to the current status of transgenic regulations.

For chemical and irradiation treatments, the dose of mutagen is an important consideration. A low mutagen dose will require a larger population to recover phenotypes of interest. In contrast, high dosage treatments will affect plant viability and fewer individuals will be recovered. If possible, a researcher may wish to test varying mutagen doses on subsets of seeds to identify the desirable dose to use on the whole population. Alternatively, the founder seed batch can be divided into several subsets with each receiving a different dose of the mutagen and subsequently handled separately. Approximately 10-15% of the founder line seed should not be mutagenized, so that this seed can be used as a wild-type reference during the M_1 , M_2 , and subsequent phenotypic screenings. It is worth noting that if a mutation of interest is identified in a heavily mutagenized line, several rounds of backcrossing into a wild-type line, to remove background mutations, may be required in order to accurately characterize the mutant phenotype.

3. Determine the size of the M_1 and M_2 populations

The chance of recovering a desired mutation is a function of the number of M_1 plants grown. M_1 plants grown from mutagenized M_1 seeds will be chimeric, and thus self-pollinating the M_1 plants is required to generate non-chimeric M_2 plants. When advancing the population, one must consider if one or more M_2 individuals should be screened per M_1 plant. Rédei et al. (1984) provides a more thorough review of these

considerations. Basically, the researcher must balance the cost and effort to generate M_2 seed from different M_1 plants (mutating M_1 seeds, growing, and harvesting more M_1 plants) against the effort to screen more M_2 plants per M_1 plant. If long term seed storage is an option, a researcher may consider screening only a portion of the M_2 population at any one time in order to decrease the number of plants assayed per screening.

4. Select a M_1 to M_2 Advancement procedure

Considerations should be made with regard to either bulk or pedigree advancement of the M_1 individuals. As previously mentioned, if the material is to be used for a specific screening purpose with strong selection (e.g. tolerance to herbicide treatment) bulk-harvesting M_1 plants could save considerable time and effort with little downside. Alternatively, if the goal is to develop a mutant population resource for future mutant screening, pedigree advancement of M_1 individuals is preferred. Alternatively, Lighner and Caspar (1998) describe a method of pooling multiple M_1 plants together into small bulks and generating multiple small bulks for the whole population. The authors suggested this pooling method as a way to reduce the work of advancing tens of thousands of M_1 individuals required to identify a mutant for a trait that is expected to have an extremely low frequency of occurrence.

5. Screen and harvest M₂ individuals

The screening methods utilized for a mutant population will vary greatly based on the objectives of the project, the design of the population, and the resources available. However, some rough guidelines are applicable across projects and designs. For visual field screening of mutants, it is helpful to have repeated check rows of the founder line to serve as a visual reference of the wild-type phenotype. Depending on the trait of interest, screening may need to occur at different locations and at multiple times during the season. For example, subsets of a population may be planted at staggered times in a field with sandy soils if root phenotyping is desired, as this will extend the period of phenotyping and facilitate the time-consuming task of digging up plants to observe roots. For developing mutant populations as a genetic resource for future screenings, it is advisable to tag, tissue sample, and individually harvest each M₂ plant. DNA from this leaf tissue can be used for reverse genetic screens.

6. Advance future generations

The methods used to advance future generations will also be determined by the goals of the project as well as the resources available. For each mutant, it is important to utilize and maintain a clear naming system that can be easily understood and traced back through previous generations. Single seed descent is a good method to use for advancing soybean mutant populations. However, if a heterozygous M₂ or later generation individual produces a segregating row, multiple mutant and wild-type individuals can be

harvested from that row in order to maintain segregating families that can be used to map the trait. An easy naming method to distinguish between mutant and wild-type individuals from a segregating row may involve adding a simple suffix of an 'A', 'B', or 'C' to the name of mutant individuals and 'Z', 'Y', or 'X' to the name of wild-type individuals. The use of barcoded tags can also facilitate the organization and maintenance of a mutant population. In order to validate quantitative traits such as seed composition, which can vary by year and location, seed composition analysis can be conducted on multiple individually harvested plants per line and on many (over 100) founder line control plants from the same field. Lastly, be prepared to take advantage of unexpected phenotyping opportunities, such as recording differences in insect herbivory or disease resistance during a particularly strong infestation and/or outbreak.

7. Long term seed storage

When generating a mutant population that is intended to be a long-term genetic resource for the community, it is important to maintain high seed viability in addition to maintaining a well-organized seed stock. Soybean seed requires frequent planting which can be burdensome with large numbers of mutant lines. To simplify the planting out of the mutant lines, the population could be divided into two groups, alternating the plantings between the two groups from year to year. Alternatively, a core collection subset of mutant lines possessing the greatest genetic diversity can be developed, concentrating the efforts for maintaining, phenotyping, and genotyping to just the most interesting or promising lines.

Currently, with the exception of the USDA Type and Isoline Collections, long-term maintenance of soybean mutant populations are the responsibility of the individual research groups that develop and use the materials. However, there is a great deal of interest within the soybean breeding and genetics community to eventually develop a public mutant repository that can maintain these valuable resources in perpetuity.

CONCLUSIONS AND FUTURE PROSPECTS

This paper has focused on the currently available soybean genetic resources. It is intriguing to speculate on how these resources will be leveraged by research programs in the future and what new tools and resources may be developed. There appears to be growing interest in the research community for exploring the value of alleles from PIs, *G. soja*, and tertiary *Glycine* accessions not currently found in *G. max* elite germplasm. It is also intriguing to speculate on the future utilization of induced genetic variation in the forms of chemical, irradiation, and transposon mutagenesis, and also transgene insertion and site specific nucleases.

While the aforementioned genetic resources can be effectively utilized on their own, one can also imagine powerful ways to combine these resources. For example, a researcher may initially screen a mutant population and eventually map and clone a gene important for abiotic stress tolerance. The researcher could mine *G. max* or *G. soja* sequence databases to identify lines with different alleles for that gene and then test the effects of a whole allelic series. Finally, the researcher could use genetic transformation

to place different alleles of that gene into isogenic backgrounds or design new alleles using SSNs. Never before have soybean researchers had so many genetic resources for gene discovery and crop improvement, and at the same time, never before have soybean researchers had such a responsibility to wisely leverage these genetic resources to sustainably improve soybean yield and quality.

CHAPTER 2

Identical Substitutions in Magnesium Chelatase Paralogs Result in Chlorophyll-Deficient Soybean Mutants

SUMMARY

The soybean [*Glycine max* (L.) Merr.] chlorophyll-deficient line MinnGold is a spontaneous mutant characterized by yellow foliage. Map-based cloning and transgenic complementation revealed that the mutant phenotype is caused by a nonsynonymous nucleotide substitution in the third exon of a Mg-chelatase subunit gene (*Chl1a*) on chromosome 13. This gene was selected as a candidate for a different yellow foliage mutant, T219H (*Y11y11*), that had been previously mapped to chromosome 13. Although the phenotypes of MinnGold and T219H are clearly distinct, sequencing of *Chl1a* in T219H identified a different nonsynonymous mutation in the third exon, only six base pairs from the MinnGold mutation. This information, along with previously published allelic tests, were used to identify and clone a third yellow foliage mutation, CD-5, which was previously mapped to chromosome 15. This mutation was identified in the *Chl1b* gene, a paralog of *Chl1a*. Sequencing of the *Chl1b* allele in CD-5 identified a nonsynonymous substitution in the third exon that confers an identical amino acid change as the T219H substitution at *Chl1a*. Protein sequence alignments of the two Mg-chelatase subunits indicated that the sites of amino acid modification in MinnGold, T219H, and CD-5 are highly conserved among photosynthetic species. These results suggest that amino acid alterations in this critical domain may create competitive inhibitory interactions between the mutant and wild-type *Chl1a* and *Chl1b* proteins.

INTRODUCTION

The soybean genetics community has an extensive history of identifying and collecting spontaneous mutant lines (Bernard, 1975; Bernard et al., 1991). Differences in leaf chlorophyll content (SOY:0001858, SOY:0001859), specifically chlorophyll-deficient phenotypes, are one of the most extensively collected classes, with more than 20 different mutants encoded by the nuclear genome identified and phenotypically characterized (Palmer et al., 2004). These mutants exhibit a diverse range of phenotypes, including dominant and recessive alleles and minor, major, intermediate, and mosaic yellow foliage types. Universally, there is a loss of plant vigor associated with the yellow foliage phenotypes, with some mutations being more detrimental to overall fitness than others. Although some of these mutations have been genetically mapped (Palmer et al., 1989, 1990; Zou et al., 2003; Mahama and Palmer, 2003; Kato and Palmer, 2004; Zhang et al., 2011), to our knowledge no study has definitively identified the causative gene or nucleotide polymorphism underlying the phenotype.

The challenge of identifying gene functions in soybean is further complicated by the high rate of gene duplication in the paleopolyploid genome (Schmutz et al., 2010). Duplicate gene copies are thought to often mask the phenotypic consequences of mutating or silencing genes. The most recent whole-genome duplication in soybean was estimated to have occurred approximately 13 million years ago, and at least two copies are maintained for nearly 75% of the genes (Schmutz et al., 2010). Therefore, mutations derived de novo in these genes may be less likely to generate new phenotypes, as the duplicate copies may mask the loss of gene functions.

The recent development of genomic and mapping tools in the soybean community has facilitated map-based cloning efforts for specific traits in recent years (e.g., Liu et al., 2012; Tian et al., 2010; reviewed by Stupar and Specht, 2013). In this study, we have used modern genomic tools to rapidly map and fine-map a recently discovered spontaneous chlorophyll-deficient mutant line known as MinnGold. The reference soybean genome sequence (Schmutz et al., 2010) was used to identify a Mg-chelatase subunit ChIIa homolog as a candidate gene. The ChII Mg-chelatase subunit is involved in catalyzing the insertion of Mg²⁺ into the protoporphyrin IX to form the first committed step in the chlorophyll biosynthesis pathway. Sequencing and functional analyses revealed that a specific nucleotide substitution in the coding region of this gene is likely responsible for the chlorophyll-deficient phenotype of MinnGold.

Knowledge gained from the cloning of the ChII Mg-chelatase gene was then used to identify the causal mutations underlying *y11* and CD-5, two chlorophyll-deficient alleles that were initially identified decades ago (Weber and Weiss, 1959; Palmer et al., 1989). A distinct nucleotide substitution six base pairs from the MinnGold mutation was identified in *y11*, consistent with previous mapping of this locus (Mahama and Palmer, 2003). Furthermore, allelism results of *y11* and CD-5 (Palmer et al., 1989) provided evidence for a Mg-chelatase paralog on chromosome 15 as a candidate gene for the CD-5 chlorophyll-deficient phenotype. Sequencing of the Mg-chelatase gene on chromosome 15 identified a nonsynonymous single-nucleotide polymorphism (SNP) change also in the third exon. Interestingly, the positions and base changes of the *y11* and CD-5 nonsynonymous polymorphisms result in identical amino acid substitutions.

MATERIALS AND METHODS

Discovery and phenotype analysis of the MinnGold mutant

In 2008, several chlorophyll-deficient plants were identified in a segregating row at a University of Minnesota winter nursery in Chile. These mutants were observed in a F₃ population derived from a cross between M99-274166 X ‘MN0091’. ‘MN0091’ is a cultivar release from the University of Minnesota Soybean Breeding Program, and M99-274166 is a selection from the cross PI 548379 X (Mandarin Ottawa X NK S19-90). Each of the chlorophyll-deficient plants was threshed and maintained as separate lines, and one of these lines was given the name MinnGold. To determine chlorophyll content, tissue was collected from the second true leaves at the V5 leaf stage of several MinnGold and ‘Williams 82’ plants and frozen in foil packets shortly after collection. Five 1-cm leaf disks were taken from frozen tissue of each cultivar. Disks were ground in cold methanol and centrifuged according to Porra et al., (1989), and absorbance was determined following the methods described by Lichtenthaler (1987) and Wellburn (1994).

Genetic mapping of chlorophyll deficiency in MinnGold

The chlorophyll-deficient mutant line MinnGold was crossed to soybean accessions ‘Archer’ and ‘Minsoy’ to develop populations for genetic mapping experiments. The F₂ progeny were grown in standard greenhouse conditions and visually phenotyped for

relative chlorophyll content. Bulk segregant analysis (BSA) (Michelmore et al., 1991) was conducted on F_2 individuals using the Illumina soybean 1536 SNP chip for genotyping (Hyten et al., 2008, 2009). A DNeasy Plant Mini Kit (QIAGEN) was used for the DNA extraction from the fresh bulk leaf tissue and for all subsequent DNA extractions from fresh leaf tissue unless otherwise stated. The wild-type and mutant bulks for both populations were composed of leaf tissue from 20 green and 20 yellow plants, respectively. BSA results from the two populations were combined to identify the approximate chromosomal position of the causative mutation.

Seeds were harvested from each F_2 individual to generate $F_{2:3}$ families, which were subsequently planted in the greenhouse. Leaf tissues were collected from one individual in each $F_{2:3}$ family, and DNA was extracted using a BioSprint 96 DNA Plant Kit (QIAGEN) following the manufacturer's protocol. To narrow the mapped interval, the $F_{2:3}$ individuals were genotyped at the University of Minnesota Genomics Center (UMGC) core facility using a custom panel of 30 SNP assays on a Sequenom MassARRAY genotyping platform. Five F_3 individuals were identified as heterozygous in the mapped interval, and these individuals were harvested and advanced to generate segregating $F_{3:4}$ families. $F_{3:4}$ families were grown in the greenhouse and leaf tissue was collected from each individual for DNA extractions, as described above. The $F_{3:4}$ individuals were subjected to an additional round of fine-mapping on a second custom Sequenom MassARRAY genotyping platform of 70 SNPs to further narrow the mapped interval. The Sequenom MassARRAY SNP panels for both rounds of fine mapping were designed based on assays from the SoySNP50K set (Song et al., 2013). The fine-mapped

interval included gene model Glyma13g30560 (renamed as Glyma.13G232500 in the Glyma.Wm82.a2.v1 annotation), a putative Mg-chelatase subunit Chl1a.

Identification of Glyma13g30560 as a candidate for *y11*

Mahama and Palmer (2003) conducted linkage mapping of multiple phenotypic traits on chromosome 13 and established the locus order: *y11*, a semidominant yellow foliage mutation; *w1*, a white flower mutation; and *y23*, a yellow foliage mutation. Their mapping data suggested that *w1* and *y23* were more closely linked and that *y11* was more distantly linked to *w1*. The combination of the mapping data from Mahama and Palmer (2003) and Zabala and Vodkin (2007) indicate that the *w1* locus and *y23* are on the opposite chromosome arm as the F_{3:4} fine-mapped interval, indicating that the *y11* and MinnGold mutations are likely on the same chromosome arm. Furthermore, the mapping results by Mahama and Palmer (2003) did not rule out the possibility that the mutation in MinnGold could be allelic to *y11*.

Sequencing of Glyma13g30560 and Glyma15g08680 in mutant and wild-type lines

The candidate gene for the MinnGold phenotype, Glyma13g30560, was sequenced from a homozygous mutant and a homozygous wild-type plant to identify the presence of polymorphisms. Homozygous individuals were identified as individuals of families not segregating for the foliage phenotype. The RNA was extracted from fresh leaf tissue using a QIAGEN RNAeasy kit following the manufacturer's protocol. The RNA was

reverse transcribed using SuperScript III reverse transcriptase (Invitrogen) following the manufacturer's protocol. The cDNA sequence for Glyma13g30560 was amplified with polymerase chain reaction (PCR) from the plants using primers 5' TACAGTCTGTCTTCTCTTCTCTTCTCCGG 3' and 5' GAATACAAACCGTGTTACATCTATGATCC 3'. The PCR amplicons from this reaction, and subsequent PCRs unless otherwise stated, were purified using a QIAquick PCR Purification Kit (QIAGEN) following the manufacturer's protocol and sequenced using the Sanger sequencing method at the UMGC.

The candidate genes for *y11* and for CD-5, Glyma13g30560 and Glyma15g08680 (renamed as Glyma.15g080200 in the Glyma.Wm82. a2.v1 annotation), respectively, also were sequenced to identify the presence of polymorphisms. The DNA was extracted from fresh leaf tissue of two *Y11/Y11* and two *y11/y11* individuals. Glyma13g30560 was amplified with PCR from the four individuals using a set of PCRs with overlapping amplicons (Table 3). Several long PCRs with a unique primer outside of the gene were needed to amplify the internal sequence of Glyma13g30560, which is highly conserved between the four soybean homologs. To obtain quality sequences for the internal sections of the gene, these long PCR amplicons were gel extracted using a QIAquick Gel Extraction Kit (QIAGEN) following the manufacturer's protocol and were used as the template for a second round of PCR using primers that anneal to the internal sequences of Glyma13g30560. This method prevented the contamination of the Glyma13g30560 sequences with sequences from the other soybean homologs. All PCRs were conducted using the proof reading KOD DNA polymerase enzyme (Novagen). The resulting PCR

amplicons were then purified. The same steps of PCR amplification, gel extraction, a second round of PCR, and PCR purification were also used to sequence Glyma15g08680 in two homozygous wild-type and two CD-5/CD-5 individuals (primers in Table 4). The final PCR amplicons were then sequenced. Sequences were aligned using Mega 5 software (Tamura et al., 2011).

To test for the occurrence of the *y11* and MinnGold candidate SNPs in a diverse set of the soybean germplasm, the portion of the third exon of Glyma13g30560 containing the candidate SNPs was PCR amplified from 29 diverse parent lines (McHale et al., 2012) using primers 5'GGCCAGGCCTTTGCATTTTG 3' and 5'ACTCAGCACACACCTTGGAG 3'. To test for the occurrence of the CD-5 candidate SNP in the soybean germplasm, the portion of the third exon of Glyma15g08680 containing the SNP was also PCR amplified from these 29 diverse parent lines using primers 5'GGCTAGGCCTTTGTGTTTGA 3' and 5'AACGGGAAATGCTGATTGAG 3'. The resulting PCR amplicons were then purified and sequenced. Sequences were aligned using Mega 5 software (Tamura et al., 2011).

CAPS assays for *y11* and CD-5

The candidate *y11* SNP change creates a SacII digestion site that can be easily screened by a cleaved amplified polymorphic sequences (CAPS) assay (Konieczny and Ausubel, 1993). Thus a CAPS assay was used to genotype additional individuals segregating for *y11* to test if the candidate SNP cosegregated with the chlorophyll deficient phenotype. Genomic DNA was extracted from fresh leaf tissue from 19

segregating progeny derived from a single heterozygous (*Y11/y11*) plant. The portion of the third exon containing the candidate SNP was PCR amplified using primers 5'GGCCAGGCCTTTGCATTTTG 3' and 5'ACTCAGCACACACC TTGGAG 3'. The PCR was then digested with *SacII* (New England BioLabs) following the manufacturer's protocol and run on a 1% agarose gel. The *SacII* digestion cuts the mutant strand forming band sizes of 237 bp and 691 bp.

The same techniques were also used to conduct a CAPS assay for CD-5. In summary, the candidate CD-5 SNP change removes an *AlwNI* digestion site that can be easily screened by a CAPS assay. As before, a CAPS assay was used to genotype 17 segregating individuals derived from a single heterozygous (wild-type/CD-5) plant to test for cosegregation of the candidate SNP with the chlorophyll deficient phenotype. Genomic DNA was extracted from fresh leaf tissue, and the portion of the third exon containing the candidate SNP was PCR amplified using primers 5'GGCTAGGCCTTTGTGTTTGA 3' and 5'AACGGGAAATGCTGATTGAG 3'. The PCR was then digested with *AlwNI* (New England BioLabs) following the manufacturer's protocol and run on a 1% agarose gel. The *AlwNI* digestion cuts the wild-type strand forming band sizes of 695 bp and 803 bp.

Transformation and transgene analysis

MinnGold was transformed with the wild-type allele of Glyma13g30560 for complementation analysis. Glyma13g30560 was amplified with PCR from wild-type cultivar 'Williams 82' using a proof reading KOD DNA polymerase enzyme (Novagen)

and the following forward and reverse primers:

5'GCTCACATGCGCGGCCGCTGGCACCCACTA ACATTTCC 3' and

5'GCTCTCATGCCCTGCAGGCGAGGAAAGA GAATGGATGG 3', respectively.

The PCR amplified 4416 bp, spanning the region 777 bp upstream of the 5' untranslated region (UTR) to 965 bp downstream of the 3'UTR. The forward and reverse primers were designed with *NotI* and *SbfI* restriction sites, respectively. The PCR fragment was gel extracted using the QIAquick Gel Extraction Kit (QIAGEN) following the manufacturer's protocol, and the fragment was cloned into the pCR-Blunt II-TOPO vector (Invitrogen) following the manufacturer's protocol. Insertion of the clone into the pCR-Blunt II-TOPO vector was confirmed using an internal primer

5'GCACCTTCAAGCTTCGCTTT 3' and M13 Forward and M13 Reverse primers

5'GTAAAACGACGGCCAG 3' and 5'CAGGAAACAGCTATGA 3', respectively.

Next, the fragment was transferred from the pCR-Blunt II-TOPO vector to the binary vector pNB96 (Fusaro et al., 2006) using a restriction digest with *NotI* and *SbfI* followed by ligation. The resulting clone was sequenced at the UMGC. The construct was transformed into MinnGold following the whole plant transformation and herbicide selection protocol described by Curtin et al. (2011). DNA was extracted from fresh T₀ leaf tissue, and the whole plant transformation was confirmed by Southern Blot analysis of both the MinnGold parent (negative control) and the transformed T₀ MinnGold plant.

The green T₀ plant was self-pollinated to produce segregating T₁ progeny. To test whether the transgene cosegregated with the wildtype (green) phenotype, DNA was extracted from fresh T₁ leaf tissue for PCR analysis. PCR primers 5'

AAGATGTTTCTCCCCCATCC 3' and 5' CGTCTTGATGAGACCTGCTG 3' were used to screen for the presence of the transgene in seventeen of the segregating T₁ progeny. The forward and reverse primers were designed to amplify a 127-bp fragment spanning from the end of the Glyam13g30560 genomic clone to the beginning of the CaMV35S promoter that precedes the BAR gene in the construct. Primers 5' GAGCTATGAATTGCCTGATGG 3' and 5' CGTTTCATGAATTCCAGTAGC 3' were used to amplify a 118-bp fragment of the soybean actin gene Glyma15g05570 as a positive PCR control.

Analysis of sequence similarity between the soybean ChII subunits

To compare the similarity of the four soybean ChII subunits, a unweighted pair group method with arithmetic mean neighbor joining tree was constructed in MEGA5 by the unweighted pair group method with arithmetic mean method (Sneath and Sokal, 1973) using amino acid sequences. The Poisson correction method was used to compute the evolutionary distances, which were calculated in units of the number of amino acid substitutions per site (Zuckerandl and Pauling, 1965). Positions containing gaps and missing data were eliminated, and a total of 415 positions were used in the final dataset. MEGA5 was used to conduct the evolutionary analyses (Tamura et al., 2011).

RESULTS

Identification, mapping, and fine-mapping of the MinnGold mutation

The spontaneous, nonlethal yellow foliage mutant (Figure 1) was first observed in a segregating F₃ population derived from a cross between M99-274166 X ‘MN0091’. The F₃ population was advanced by modified single seed descent and exhibited a segregation ratio of 157 green (wildtype) plants to 50 yellow (mutant) individuals. The ratio of 157:50 for the wild-type:chlorophyll-deficient phenotype is consistent with a singlelocus, recessive mutation (χ^2 test, $p = 0.78$). The mutant phenotype appears more vigorous than previously identified chlorophyll-deficient soybean mutants. Measurement of leaf chlorophyll levels from tissue collected from the second true leaves at the V5 leaf stage show that this mutant displayed a dramatic reduction of both chlorophyll a and chlorophyll b relative to wild-type soybean plants (Figure 1). This mutant was given the name MinnGold because of its bright yellow leaf coloration during early foliar development.

MinnGold was crossed to two soybean accessions, ‘Archer’ and ‘Minsoy’, to develop mapping populations. The F₂ progeny from both populations exhibited segregation ratios of 3:1 for the wild-type: chlorophyll-deficient phenotype. The yellow foliage phenotype was observed in 79 of 311 plants (χ^2 test, $p = 0.87$) derived from the MinnGold X ‘Archer’ cross. The yellow foliage phenotype was observed in 48 of 186 plants (χ^2 test, $p = 0.80$) derived from the MinnGold X ‘Minsoy’ cross. These results further indicate that the phenotype is caused by a recessive allele at a single, nuclear

locus. BSA (Michelmore et al., 1991) was conducted on wild-type and mutant F₂ bulks from both populations. The BSA populations were genotyped using the Golden Gate 1536 SNP chip assay (Hyten et al., 2009) to narrow the region of interest to a 14.3 Mb interval on chromosome 13 (Figure 2).

Genotyping of 372 F₃ individuals and parental lines was subsequently performed using a customized panel of 30 SNP markers within this chromosome 13 region, reducing the interval to 2.9 Mb. This interval was flanked by markers at positions 32,120,834 and 35,053,712 based on the soybean genome assembly version Glyma. Wm82.a1 (Figure 2). Five green F₃ individuals that were genotyped as potentially heterozygous at or within either flanking marker were also found to segregate for the mutant phenotype in the F_{3:4} generation. In total, 159 F_{3:4} individuals from these families, along with a subset of F₃ individuals and parental line controls, were phenotyped and genotyped with a custom panel of 70 SNP markers. After this round of mapping, the mutation was mapped to a 165.3-kb interval on chromosome 13, residing between positions 33,141,206 and 33,306,556 (Figure 2).

Identification of the candidate gene

According to the soybean genome assembly version 1.1 (Schmutz et al., 2010), the 163.5-kb interval contains 22 annotated genes. This list (Table 5) includes Glyma13g30560, a gene encoding a putative Mg-chelatase subunit Chl1a. This gene was a promising candidate for the mutant phenotype since Mg-chelatase is involved with a major step in chlorophyll biosynthesis. Additionally, previous reports in thale cress

(*Arabidopsis thaliana*) (Mochizuki et al., 2001; Rissler et al., 2002; Kobayashi et al., 2008; Huang and Li, 2009), rice (*Oryza sativa*) (Zhang et al., 2006; Jung et al., 2003), barley (*Hordeum vulgare*) (Hansson et al., 1999; Hansson et al., 2002; Jensen et al., 1996), maize (*Zea mays*) (Sawers et al., 2006), and tobacco (*Nicotiana tabacum*) (Nguyen, 1995) have associated mutations in Mg-chelatase subunits with chlorophyll deficient phenotypes.

The candidate gene Glyma13g30560 was sequenced in MinnGold and compared with the reference genome sequence (Schmutz et al., 2010). A single, nonsynonymous substitution identified in the third exon results in an amino acid substitution of Arginine to Glutamine (R273Q) (Figure 3). A comparative sequence analysis of Glyma13g30560 across twenty-nine diverse soybean lines (McHale et al., 2012) found that the single base substitution was unique to MinnGold. All 29 of the diverse lines exhibited the wild-type sequence (Figure 3). The spontaneous occurrence of the mutant phenotype and the unique nature of this specific mutation suggest that this allele appeared de novo within the breeding population.

Transgenic validation of the ChII1a Mg chelatase subunit

The full Glyma13g30560 gene model was cloned from a PCR product amplified from the wild-type cv. ‘Williams 82’. The construct, which contained the genomic sequence spanning from 777 bp upstream of the 5’UTR to 965 bp downstream of the 3’UTR, was transformed into MinnGold to test for phenotypic complementation. A successfully transformed T₀ plant displayed the wild-type green foliage phenotype.

Southern blot analysis indicated that this plant harbored a single transgenic locus (data not shown). The subsequent T₁ progeny segregated in a ratio of 39 green plants to 11 yellow plants, following a 3:1 segregation ratio consistent with a single transgenic locus (χ^2 test, $p = 0.62$). PCR analysis of the segregating progeny found that the transgenic locus perfectly cosegregated with the green phenotype (Figure 4). This transgenic complementation further indicates that the mutation in the coding region of Glyma13g30560 is responsible for the MinnGold phenotype. Chlorophyll levels were measured in the second true leaf at the V5 growth stage of several MinnGold, ‘Williams 82’, and segregating T₁ progeny to quantitatively assess the phenotypic effect of the transgene. The chlorophyll levels of wild-type ‘Williams 82’ and transgenic MinnGold plants were not found to be different ($p = 0.26$), whereas the transgenic MinnGold individuals had significantly greater chlorophyll than the untransformed MinnGold line ($p < 0.0001$), thus indicating that the transgene successfully recovered the wild-type phenotype (Figure 1B).

Sequence analysis suggests the mutation in MinnGold is allelic to *y11*

The combination of mapping results by Mahama and Palmer (2003) and Zabala and Vodkin (2007) revealed the coincidental mapping of *y11*, another chlorophyll deficient mutant (Figure 5A) to the same chromosome arm as the mutation in MinnGold, and led us to consider that *y11* may also be caused by a mutation in Glyma13g30560. Homozygous mutant (*y11/y11*) plants of the mutant T219H can grow through the seedling stage, but do not survive to set seed. Therefore, a family derived from a single

heterozygous T219H plant (*Y11/y11*) was grown to collect DNA from homozygous mutant individuals. Sequencing of Glyma13g30560 from two homozygous chlorophyll deficient yellow mutants (*y11/y11*) revealed a single nonsynonymous SNP change of adenine to guanine in the third exon of Glyma13g30560 [resulting in a change of glutamine to arginine (Q275R)], six base pairs downstream from the SNP change in MinnGold (Figure 3). Glyma13g30560 sequence from two homozygous dark green individuals (*Y11/Y11*) derived from the same segregating T219H family indicated that both of these individuals had the wild-type sequence. A CAPS assay of nineteen individuals segregating for the presence of the candidate *y11* SNP found that the SNP perfectly cosegregated with the foliage phenotype (Figure 6). These results indicate that the mutation in MinnGold would be allelic to *y11*. Therefore, the causative mutation in MinnGold has been assigned the trait designation *y11-2*.

Identification of identical yellow foliage mutations at paralogous magnesium chelatase genes

Palmer et al. (1989) found that the *y11* and CD-5 chlorophyll deficient mutants display nearly identical chlorophyll deficient phenotypes (Figure 5). Allelism test results and similarity in phenotype led the authors (Palmer et al., 1989) to initially consider that the two mutants were allelic and thus were surprised to find that CD-5 cosegregated with the blunt pubescence tip locus (*pb*) on chromosome 15. After identifying the likely *y11* causative mutation in a Mg-chelatase subunit, we identified a Mg-chelatase paralog on chromosome 15, Glyma15g08680, as a candidate gene for the CD-5. Sequencing of

Glyma15g08680 from two homozygous CD-5 individuals and from two wild-type individuals all derived from a segregating family identified a single non-synonymous SNP in the third exon. A CAPS assay of seventeen individuals segregating for the presence of CD-5 showed perfect cosegregation of the candidate SNP with the mutant phenotype (Figure 7). These results suggest that the identified SNP is causative of the CD-5 chlorophyll-deficient phenotype. A comparative sequence analysis across 29 diverse soybean lines (McHale et al. 2012) found that the single base substitution in Glyma15g08680 was unique to CD-5. All 29 of the diverse lines exhibited the wild-type sequence (Figure 8A). Remarkably, the Glyma15g08680 amino acid change in CD-5 (Q275R) was identical in position and sequence to the Glyma13g30560 change in *y11* (Q275R) (Figure 8B and Figure 9).

Sequence comparison of the Mg-chelatase subunit across photosynthetic species

The clustering of the altered amino acids in the *y11*, *y11-2*, and CD-5 mutants suggests that the affected residues are located in a domain critical for protein function. A comparison of the amino acid sequences of Mg-chelatase ChII subunit revealed remarkable conservation across angiosperm species (Figure 9). Furthermore, the three nonsynonymous substitutions found in *y11*, *y11-2*, and CD-5 occur at residues that are completely conserved across diverse photosynthetic species, suggesting that these positions are critical for proper ChII function.

DISCUSSION

Mg-chelatase is a protein complex comprising three subunits (ChII, ChID, and ChIH) that catalyze the insertion of Mg^{2+} into the protoporphyrin IX as the first committed step in the chlorophyll biosynthesis pathway. These three subunits are highly conserved and have been found in all photosynthetic organisms. Both in vitro and in vivo evidence from bacteria to higher plants have shown that all three components are essential for proper Mg-chelatase function (Gibson et al., 1995; Hansson et al., 1999, 2002). In soybean, earlier work on the Mg-chelatase ChII subunit has been limited to molecular characterization, including sequence description, cellular localization, and expression of the putative ChII (Nakayama et al., 1995).

The data presented here show that the T219H and MinnGold chlorophyll-deficient phenotypes perfectly cosegregate with nonsynonymous SNPs located in the third exon of the gene model Glyma13g30560, a Mg-chelatase subunit ChII1a homolog. We also show that the CD-5 chlorophyll-deficient phenotype perfectly cosegregates with a nonsynonymous SNP located in the third exon of the gene model Glyma15g08680, a Mg-chelatase subunit ChII1b homolog. The altered residues in *y11*, *y11-2*, and CD-5 are in positions that are conserved across a diverse panel of photosynthetic species, as well as a diverse panel of soybean accessions. Furthermore, genetic transformation of the wild-type Glyma13g30560 allele into the MinnGold background recovered the green foliage (wild-type) phenotype. Collectively, these data provide sufficient evidence to conclude that the chlorophyll deficiency phenotypes of T219H, MinnGold, and CD-5 are caused by mutations in paralogous genes encoding Mg-chelatase subunit ChII1a (*y11* and *y11-2*)

and ChII1b (CD-5).

In addition to the similar phenotypes exhibited by the *y11* and CD-5 mutants, it was found that these alleles have identical amino acid substitutions (Q275) that occur in the different paralogs. The mutations occurred in two different lines (T219H and CD-5) rather than in a single line, suggesting that gene conversion was not causative of the mutation. To our knowledge, this is the first time that identical mutations have been identified in different paralogs in two phenotypically similar mutant plants.

A high level of residue conservation is seen across species for the Mg-Chelatase subunit as a whole, and the complete amino acid conservation observed specifically at the region in which the *y11*, *y11-2*, and CD-5 mutations occur suggests a high degree of specificity is required for proper functioning (Figure 9). Even the four putative Mg-Chelatase ChII subunit paralogs of soybean share identical amino acid residues at the mutated positions (data not shown). The amino acid substitutions of *y11* and *y11-2* are separated by only two residues (Q275R and R273Q, respectively), suggesting that these mutations may affect a similar domain in the ChII1a subunit (Figure 9). The occurrence of the CD-5 mutation, also in this interval but in a separate paralog, further indicates that mutations in this interval disrupt a domain critical for ChII function (Figure 9 and Figure 10). The *y11*/CD-5 and *y11-2* mutations are seven and nine residues downstream of a predicted ATP binding domain, respectively, and four and six residues upstream of a predicted Arginine finger domain, respectively (Marchler-Bauer et al., 2013). Additionally it is predicted that the *y11-2* mutation occurs at a residue immediately preceding an alpha helix domain, while the *y11*/CD-5 mutation occurs within an alpha

helix domain (Kelley and Sternberg, 2009). However, it is not yet known if these mutations disrupt the ATP binding domain and/or the Arginine finger domain.

The phenotypes of *y11* and CD-5 are different than *y11-2*, despite being caused by nonsynonymous mutations that are only two amino acid residues apart (Figure 9 and Figure 11), suggesting that there is a high level of specificity required for proper interaction and function of the Mg-chelatase subunits. Interestingly, the *y11* and CD-5 alleles are semidominant whereas the *y11-2* allele is completely recessive. The semidominance of the *y11* and CD-5 alleles follows what has been observed in other species, where missense ChII mutations have resulted in semidominant phenotypes (Nguyen, 1995; Jensen et al., 1996, Hansson et al., 1999, 2002; Soldatova et al., 2005; Sawers et al., 2006; Huang and Li, 2009). In contrast, previous studies have found that completely recessive ChII alleles tend to be caused by presumably more detrimental molecular alterations that truncate the gene or decrease expression (Fischerova, 1975; Koncz et al., 1990; Rissler et al., 2002; Sawers et al., 2006; Huang and Li, 2009; Kobayashi et al., 2008). This appears to be contradictory, as the stronger phenotypic effect (semidominant as opposed to recessive) is associated with a presumably less influential alteration (e.g., an in-frame base substitution as opposed to a gene knockout).

The semidominant nature of the missense alleles is predicted to be a result of inhibitory interactions between the mutant and wildtype ChII subunits. Fodje et al. (2001) found that the *R. capsulatus* Mg-chelatase subunit orthologous to ChII1 assembles into a hexameric ring structure complex, and Hansson et al. (2002) demonstrated that for the complex to have proper activity, each segment in the hexamer needs to be capable of

ATP hydrolysis. Hansson et al. (2002) and Sawers et al. (2006) demonstrated that the hexameric ring ATPase activity is detrimentally affected when the full-length mutant proteins are assembled into the hexameric ring with the wild-type proteins. The semidominant alleles are thus a result of the mutant ChII proteins, which do not have full ATPase activity, impeding the function of the wild-type ChII subunits in the hexameric ring. Therefore, the semidominance of the *y11* and CD-5 alleles identified in this study indeed fit this model.

The *y11-2* allele, however, does not fit this model. This allele is based on a missense mutation at nearly the same location as *y11* and CD-5 but exhibits a completely recessive phenotype. One similar finding has been reported in rice, where a *chl9* missense mutation caused a recessive phenotype; however, the position of this mutation was located 40 residues downstream of the *y11-2* mutation in a region with reduced amino acid conservation (Zhang et al. 2006) (Figure 9). Still, it is unclear why the *y11-2* allele is recessive whereas the *y11* and CD-5 alleles are semidominant as all three missense mutations occur in the same domain in the ChII subunit, change the amino acid charge, and affect conserved residues. It is possible that the *y11-2* missense mutation does not interfere with the wild-type ChII proteins. Instead, the *y11-2* mutation may impede the integration of the mutant ChIIa protein into the hexameric ring and thus the mutant ChII protein cannot inhibit the function of wild-type ChIIa and ChIIb proteins (see wild-type/*y11-2* and *y11-2/y11-2* in Figure 11A). Thus, it is suggested that the *y11-2* mutant proteins are effectively not involved in chlorophyll biosynthesis. A recessive chlorophyll deficient Arabidopsis mutant, *chli1/chli1* (SAIL_230_D11), which does not

express ChII1 (Huang and Li, 2009), displays a phenotype remarkably similar to the *y11-2* phenotype. Although the two mutations are caused by different mechanisms, the common result between the two mutants suggests that neither produces a ChII1 subunit that is involved in chlorophyll biosynthesis. As a result, both mutants produce a significantly lower level of chlorophyll than wild-type (Huang and Li, 2009). For both the *chli1/chli1* and the *y11-2* mutant, paralagous activity (by CHLI2 in Arabidopsis and ChII1b (Glyma15g08680) in soybean) can partially rescue the phenotypes, assuming neither mutant produces a ChII that interferes with CHLI2 or ChII1b function, respectively (Huang and Li, 2009).

Additional duplication of the ChII genes in soybean adds another layer of complexity to understanding the genetics and interactions of this gene family. The paleopolyploid genome has retained four copies of ChII, including two ChII2 genes (Glyma07g32550 and Glyma13g24050; Figure 12). The presence of higher-order duplicated gene pairs could provide additional opportunities for sequence evolution through nonfunctionalization, subfunctionalization, or neofunctionalization of the gene paralogs (Force et al., 1999; Lynch et al., 2001). However, sequence evolution can be constrained in proteins that are part of protein complexes (Fraser et al., 2002; Szklarczyk et al., 2008). One might hypothesize that the ChII2 gene copies may influence the ChII1 mutant phenotypes, either by masking protein malfunctions or contributing additional inhibitory interactions. However, previous work (Severin et al., 2010) found that soybean ChII1a (Glyma13g30560) and ChII1b (Glyma15g08680) are expressed at much greater levels than the ChII2a and ChII2b genes, particularly in leaf tissue (Table 6).

Therefore, it is possible that the ChII2 gene copies have minimal influence on the observed mutant phenotype.

In summary, this study has identified genetic mutations of soybean ChII1 alleles that confer chlorophyll-deficient foliage phenotypes. Identical missense substitutions at paralogous gene copies were found to confer nearly identical semi-dominant mutant phenotypes, whereas a similar missense mutation *yII-2* conferred a completely recessive phenotype. We speculate that the soybean ChIIa paralogous proteins interact with one another, and the contrasting phenotypes observed from mutations a few base pairs apart may demonstrate the high level of specificity required for these interactions. The delicate nature of this interaction, along with conserved gene function, may contribute to the high sequence conservation of this duplicate gene family. Extended more broadly, the sensitivity of paralogous gene interactions may be crucial in determining whether the duplicate copies are amenable to divergence, or recalcitrant to genetic and transcriptional alterations.

CHAPTER 3

Fast Neutron induced structural rearrangements at a soybean *NAPI* locus result in *gnarled* trichomes

SUMMARY

A soybean (*Glycine max* (L.) Merr.) *gnarled* trichome mutant, exhibiting stunted trichomes compared to wild-type, was identified in a fast neutron mutant population. Genetic mapping using whole genome sequencing based bulked segregant analysis identified a 26.6 megabase interval on chromosome 20 that co-segregates with the phenotype. Comparative genomic hybridization analysis of the mutant indicated that the chromosome 20 interval included a small structural variant within the coding region of a soybean ortholog (Glyma.20G019300) of Arabidopsis *Nck-Associated Protein 1* (*NAPI*), a regulator of actin nucleation during trichome morphogenesis. Sequence analysis of the candidate allele revealed multiple rearrangements within the coding region, including two deletions (approximately 1-2 kb each), a translocation, and an inversion. Further analyses revealed that the mutant allele perfectly co-segregated with the phenotype, and a wild-type soybean *NAPI* transgene functionally complemented an Arabidopsis *nap1* mutant. Additionally, mapping and exon sequencing of *NAPI* in a spontaneous soybean *gnarled* trichome mutant (T31) identified a frame shift mutation resulting in a truncation of the coding region. These data indicate that the soybean *NAPI* gene is essential for proper trichome development and show the utility of the soybean fast neutron population for forward genetics approaches to identifying genes.

INTRODUCTION

The plant trichome is an elongated epidermal cell that undergoes cell enlargement away from the plant surface. Trichomes develop on the surfaces of leaves, stems, petioles, and some reproductive organs (Vermeer and Peterson, 1979; Nyman, 1993). Trichomes perform many biological functions, including plant defense against insect predation, where they can affect larvae growth and insect preferences (Levin, 1973; Robbins et al., 1979; Hulburt et al., 2004), and adaptation of desert plants to drought conditions by increasing the leaf reflectance, which helps to moderate leaf temperatures without requiring increased transpiration (Ehlering and Mooney, 1978). Economically, cotton seed trichomes compose the fibers that make cotton a valuable commodity. Thus, an understanding of the genetic control of trichome development impacts both agricultural and economic productivity.

There has been considerable interest in understanding the genetic controls that underlie trichome formation, particularly in model plant systems. In *Arabidopsis*, several genes required for trichome formation have been characterized. *GLABRA1 (GL1)* and *TRANSPARENT TESTA GLABRA1 (TTG1)* are important for trichome initiation (Oppenheimer et al., 1991; Walker et al., 1999). *GLABRA2 (GL2)* controls normal trichome morphogenesis (Rerie et al., 1994), and *TRIPTYCHON (TRY)* and *CAPRICE* control the spacing pattern of trichomes across the leaf surface (Schellmann et al., 2002). *ZWICHEL (ZWI)* and *TRANSPARENT TESTA GLABRA2 (TTG2)* affect trichome branching (Oppenheimer et al., 1997; Johnson, 2002), and mutations in *CROOKED (CRK)* and *GNARLED (GRL)* cause distorted trichomes (Mathur et al., 2003; Deeks et al.,

2004; El-Assal et al., 2004). However, to our knowledge, the only previously isolated soybean gene affecting any trichome trait is the T allele, which has pleiotropic effects on the color of the trichome, hilum, and seed-coat (Woodworth, 1921; Zabala and Vodkin, 2003).

Several soybean mutagenesis platforms have been developed for functional characterization of soybean genes (reviewed by Campbell and Stupar, 2016). These platforms include chemical mutagenesis (Cooper et al., 2008; Gillman et al., 2014), transposon tagging (Palmer et al., 2008a, b; Mathieu et al., 2009; Hancock et al., 2011; Cui et al., 2013; Raval et al., 2013), and irradiation mutagenesis (Men et al., 2002; Bolon et al., 2011; Gillman et al., 2014). Chemical mutagenesis causes single nucleotide polymorphisms (SNPs), and transposon tagging results in insertion mutants (Cooper et al., 2008; Palmer et al., 2008a, b; Mathieu et al., 2009; Hancock et al., 2011; Cui et al., 2013; Raval et al., 2013; Gillman et al., 2014). In contrast to the limited mutation types caused by chemical and transposon mutagenesis, irradiation mutagenesis has been reported to induce a wide variety of mutation types, including structural rearrangements (e.g. deletions, duplications, translocations, and inversions) of varying sizes, and SNPs (Bolon et al., 2011, 2014; Belfield et al., 2012).

Previously, forward genetic approaches to identify causative mutations induced through mutagenesis was often slow and typically required initial coarse mapping followed by one or more rounds of fine-mapping to positively identify a causative variant. Genetic methods, such as Bulk Segregant Analysis (BSA) (Michelmore et al., 1991), have been developed to facilitate the coarse mapping of qualitative traits,

however, fine-mapping based on the phenotyping and genotyping of individuals from large populations is arduous. The advent of new molecular technologies has rapidly decreased the time required to physically and genetically map potential causative polymorphisms. For example, irradiation-induced mutations can be detected using array Comparative Genomic Hybridization (aCGH) or genome resequencing. The aCGH approach is useful for quickly identifying sufficiently large (≥ 2 kb) deletions and duplications, as has been demonstrated in a variety of plant species such as *Arabidopsis thaliana* (Gong et al., 2004), clementine (*Citrus clementina* Hort. Ex Tan. Cv. Clemenules) (Ríos et al., 2008), rice (*Oryza sativa*) (Bart et al., 2010), and soybean (Bolon et al., 2011, 2014). Furthermore, sequencing based genotyping of BSA samples can be used to decrease the time required to identify chromosomal loci that co-segregate with qualitative traits. This concept has been demonstrated using a range of different sequencing approaches (including sequencing of RNA, whole-genomes, and exomes) and has been demonstrated in several plant species, such as tomato *Solanum lycopersicum* (Illa-Berenguer et al., 2015), maize (Liu et al., 2012; Haase et al., 2015), rice (Takagi et al., 2013; Yang et al., 2013), barley (Mascher et al., 2014) and *Arabidopsis* (James et al., 2013; Zhang et al., 2014).

We were intrigued to test whether a combination of aCGH and Whole Genome Sequencing based Bulk Segregant Analysis (WGS-BSA) could facilitate the rapid cloning of the causative gene(s) from an irradiated soybean mutant. In this study, we report the identification of a causative mutation underlying a previously identified soybean fast neutron mutant that exhibits *gnarled* trichomes. A combination of aCGH

and WGS-BSA was used to identify a *Nck-Associated Protein 1 (NAPI)* candidate gene for this trait, and subsequent genetic and molecular analyses confirmed the essential function of this gene in trichome development.

MATERIALS AND METHODS

Populations and Phenotyping

A *gnarled* trichome mutant, R55C01 (Soybase.org mutant FN0175501), was identified in a soybean fast neutron mutant population developed at the University of Minnesota using the soybean line ‘M92-220’ which as derived from the variety ‘MN1302’ (Orf and Denny, 2004; Bolon et al., 2011). This mutant was crossed to the wild-type accession ‘Noir 1’ (subline Noir 1-SGC-01 (McHale et al., 2012)) to generate a segregating mapping population. The F₁ hybrid and the subsequent segregating F₂ and F₃ individuals were grown in the greenhouse and visually phenotyped.

Detection of structural variants using comparative genomic hybridization microarrays

The aCGH array was designed using the first version of the soybean reference cv. ‘Williams 82’ genome sequence Glyma.Wm82.a1.v1.1 (Bernard and Cremeens, 1988;

Schmutz et al., 2010). The array was composed of unique sequence probes (50-60mers) spaced across the genome at an interval typically ranging from 0.5-1.1 kb. The methods used for the labeling and the aCGH analysis were conducted according to the methods described in previous studies (Haun et al., 2011; Bolon et al., 2011, 2014; Anderson et al., 2014), using ‘M92-220’ as the reference sample for the array. Genomic DNA was isolated from leaf tissue using the Qiagen DNeasy kit, and 500 ng of genome DNA from each line was used for the labeling reaction. The mutant DNA was labeled using Cy3 dye and ‘M92-220’ reference sample was labeled with Cy5 dye. The labels were incorporated using the 3’–5’ exo-Klenow fragment from DNA polymerase I. The labeled DNA was quantified and hybridized for 72 hrs at 42° C to the 700k feature NimbleGen aCGH array. The methods used for array scanning and data analyses have been previously described (Bolon et al., 2011).

Sequencing of R55C01, ‘Noir 1’, and F₂ bulks

Fifty F₂ individuals with wild-type trichomes and fifty F₂ individuals with mutant trichomes were chosen from the ‘Noir 1’ x R55C01 population to compose the two mapping bulks. Genomic DNA of both bulks, the mutant R55C01, and the wild-type ‘Noir 1’ individual was extracted from leaf tissue using a Qiagen DNeasy kit. DNA samples were submitted to the University of Minnesota Genomics Center (UMGC) for sequencing on an Illumina HiSeq 2000 producing 101bp paired-end reads with the goal of achieving an average sequencing coverage of 30x. Scythe

(<https://github.com/vsbuffalo/scythe>) was used to remove adapter sequences from the 3' ends of reads, with a 5% prior on contamination rate. Sickle (<https://github.com/najoshi/sickle>) was then used to remove bases with a Phred quality below 20. Cleaned reads were aligned to the updated soybean reference genome assembly Glyma.Wm82.a2.v1 (Song et al. 2016) using BWA-MEM version 0.7.5a (Li, 2013). Mismatch penalties and alignment reporting parameters were adjusted to report alignments with approximately 1% mismatch from the reference. Alignments were sorted, de-duplicated, and labeled with read groups using Picard Tools version 1.107 (<http://broadinstitute.github.io/picard/>). Alignments were then re-aligned around potential insertion/deletion polymorphisms using the Genome Analysis Tool Kit (GATK) version 3.1-1. Variant positions for the SoySNP50K genotyping platform (Song et al. 2013) were called using the GATK UnifiedGenotyper (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013). A custom Python script was then used to estimate allele frequencies and calculate read depths in each bulk.

Whole Genome Sequencing based Bulked Segregant Analysis (WGS-BSA)

The bulk allele frequencies were initially calculated as the Glyma.Wm82.a2.v1 reference or the alternate state using custom PERL script for the SoySNP50K positions (Song et al., 2013; Song et al., 2016). To utilize this data for mapping, the allele frequencies were converted to allele frequencies of the wild-type 'Noir 1' parent at each position in both bulks based on the alternate or reference allele state of 'Noir 1' at each

SNP position from the ‘Noir 1’ sequence data. Non-polymorphic SNPs (having allele frequencies in both bulks of greater than or equal to 0.9 or less than or equal to 0.1), SNPs with missing data in either bulk, and SNPs with read counts less than 10 in either bulk were removed from the dataset. The allele frequencies were plotted and graphically analyzed for spreads in allele frequency.

Genotyping and Phenotyping of segregating F₃ individuals

R55C01, ‘Noir 1’, and 50 F₃ families derived from different F₂ individuals were planted in the greenhouse and visually phenotyped for the presence of wild-type or *gnarled* trichomes. Genomic DNA was extracted from a single F₃ individual from each homozygous family and from one mutant and one wild-type plant from each segregating F₃ family. The genomic sequence of R55C01 was utilized to design PCR primers (Table 7) that amplified distinct amplicons for the wild-type and mutant alleles (Figure 13).

Validation of chromosome rearrangements in *GmNAPI*

PCR was used to validate the chromosome rearrangements identified by whole genome sequencing at the locus of the candidate gene, *Glycine max NAPI* (*GmNAPI*). PCR primers were designed using the genomic sequence of R55C01 (Table 8).

RNA sequencing of R55C01 and 'M92-220' root, seed, and leaf tissue

R55C01 and 'M92-220' seeds were imbibed in sterile water for 48 hours before being transferred to pots containing quartz sand, with four seedlings planted per pot and later thinned to one plant. Plants were placed in a growth chamber at 28°C, oscillating between 16 hours of light and eight hours of dark. Each pot was watered daily with 500 mL of nutrient solution (O'Rourke et al., 2014). At the V2 stage, emerging trifoliates and total root tissues were harvested and immediately immersed in liquid nitrogen from three biological replicates. Developing seeds were harvested at seed stage 0 (10mg) from three additional biological replicates. RNA was extracted from leaf, root, and developing seeds using the Qiagen RNeasy kit and submitted to the UMGC, where samples were sequenced on an Illumina HiSeq 2000. Illumina library preparation, clustering and sequencing reagents were used throughout the process, following the manufacturer's recommendations. Samples were sequenced as 50bp paired-end reads with an insert size of 200bp. On average, each sample generated 26 million paired-end reads. Read quality was confirmed using FASTQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the reference genome version Glyma.Wm82.a1.v1 (Schmutz et al., 2010) using TopHat2 (Kim et al., 2013). Reads mapping to genic regions were identified using a combination of SAMtools and HTseq (Li et al., 2009; Anders et al., 2014). Differential gene expression and exon expression analyses were performed in R (R Development Core Team, 2006) using DESeq (Anders and Huber, 2010) and EdgeR (Robinson et al., 2010), respectively.

Mapping the *p2* introgression interval

The similarity of the R55C01 trichome mutant phenotype to the phenotypic descriptions and images of the *p2* trichomes of line T31 (PI548159) suggested that the *p2* allele was caused by a mutation affecting the same gene or pathway as the R55C01 mutant (Stewart and Wentz, 1926; Bernard and Singh, 1969; Singh et al., 1971; Healy et al., 2005). Bernard et al. (1991) backcrossed the *p2* allele into the cv. ‘Harosoy’ (PI548573) (Weiss and Stevenson, 1955) and cv. ‘Clark’ (PI548533) (Johnson, 1958) backgrounds to generate two ‘Harosoy’ *p2* backcross lines (PI547713 and PI547743) and three ‘Clark’ *p2* backcross lines (PI547449, PI547565, and PI547566). The SoySNP50K data for the three parents and five *p2* backcross lines (Song et al., 2015) was obtained from SoyBase (<http://soybase.org>). SNPs that were not polymorphic between mutant line T31 and wild-type lines ‘Clark’ and ‘Harosoy’ were removed. The genome was then scanned for SNPs at which T31 and all five *p2* backcross lines shared the same allele.

Sequencing the *GmNAPI* gene in T31

Seed of T31 (PI548159) was obtained from USDA Germplasm Resource Information Network, and T31 genomic DNA was extracted from leaf tissue. The candidate gene for this study, Glyma.20G019300, has a length of 22.5kb, not including the promoter sequence, and the predicted transcript length is 4.8 kb (Song et al., 2016). PCR primers were designed to amplify the 5’UTR, 3’UTR, and all twenty-three exons

including all splice site junctions (Table 9). Reactions were PCR purified using a QIAquick PCR Purification Kit (Qiagen) and sequenced at the UMGC. Sequences were aligned to the sequence of Glyma.20G019300 based on the reference sequence Glyma.Wm82.a2.v1.1 (Song et al., 2016).

Transgenic complementation of Arabidopsis *glr-4* using *GmNAPI*

The *GmNAPI* construct was designed in a pMDC123 backbone and consisted of the native promoter amplified from ‘Williams 82’ driving the synthesized *GmNAPI* cDNA (Piscataway, NJ) followed by a NOS terminator. The pMD123 vector used also contained a 35S promoter driving a BAR herbicide resistance gene.

The Arabidopsis *nap1* mutant, *glr-4* (El-Assal et al., 2004), was transformed with the *GmNAPI* construct using the floral dip method (Clough and Bent, 1998), and T₁ seeds were planted in a flat containing standard potting mix. Ten and seventeen days after germination the flat was sprayed with a 0.01% solution of Glufosinate and resistant plants were then transferred to individual pots. Leaf tissue was collected from each T₁ individual. PCR primers that amplified across the junction between the promoter and first exon of Glyma.20G019300 were used to test for the presence of *GmNAPI* construct in ‘Williams 82’, *glr-4*, and 20 T₁ individuals with wild-type trichomes (Table 10).

RESULTS

Identification and mapping of the *gnarled* trichome mutant

Several morphological and developmental mutants were discovered during the visual phenotypic screening of the soybean fast neutron population generated at the University of Minnesota (Bolon et al., 2011). Mutant R55C01 (Soybase mutant FN0175501) was identified as a short trichome mutant (SOY:0001804) (Figure 14A). Scanning electron microscope images of the leaves from wildtype and mutant plants indicated that the mutant has a *gnarled* trichome phenotype (Figure 14B-C), which is characterized by trichomes that are swollen, twisted and reduced in length (Szymanski et al., 1999; Deeks et al. 2004; El-Assal et al. 2004). The *gnarled* mutant trichomes are shorter, exhibit wide, flaccid shafts (SOY:0001720), lay on the surface of the leaf or stem (SOY:0001977), and have round, blunt tips (SOY:0001722).

aCGH was conducted on an M5 mutant plant to identify potential causative mutations (this plant was assigned the identification number FN0175501.x2.02.01.M5 in Bolon et al. (2011)). The aCGH results did not identify any duplications but did identify two deletions in the mutant genome: a putative 26 kb deletion on chromosome 5 and a putative ~2 kb deletion on chromosome 20. However, the aCGH method is not sensitive enough to detect some types of rearrangements (inversions and translocations), small deletions, and small duplications, which may underlie the mutant phenotype.

Genetic mapping was conducted to identify the genomic interval co-segregating with the trichome phenotype. The mutant was outcrossed to the wild-type accession ‘Noir 1’

to generate the mapping population. The F₁ plant had wild-type trichomes, and the F₂ population segregated in a 3:1 wild-type to mutant ratio (144 to 53; Chi-squared p-value = 0.537 for one locus), indicating that the trichome mutant phenotype was recessive and was caused by a mutation at a single locus.

WGS-BSA was conducted using a bulked sample of F₂ mutant segregants and a bulked sample of F₂ wild-type segregants. The allele frequencies were calculated with respect to the 'Noir 1' SNP state such that SNPs closely linked to the causative locus would exhibit a mutant bulk allele frequency of approximately zero and a wild-type bulk frequency of approximately 0.66 (as two out of every three wild-type plants would be expected to be heterozygous at the causative locus). The allele frequencies at each SNP position were visualized as the proportion of reads derived from 'Noir 1' and averaged across a 21 SNP sliding window (Figure 14D). Chromosomes 1 through 19 did not show any major divergence in allele frequencies between the bulks. However, the chromosome 20 allele frequencies exhibited the expected divergence for the causative locus (Figure 14D).

A detailed analysis of chromosome 20 showed that the mutant bulk had an average 'Noir 1' allele frequency of zero for a 26.6 Mb interval spanning the pericentromere on chromosome 20, between positions 1.67 Mb and 28.3 Mb (Figure 15A). The mutant bulk's average 'Noir 1' allele frequency of zero indicates that the mutant parent markers in the 26.6 Mb interval co-segregated with the F₂ mutant phenotype. For the same interval, the wild-type bulk had the expected average 'Noir 1' allele frequency of 0.66. The 26.6 Mb interval overlapped with one structural variant detected by the aCGH

experiment, the approximately 2 kb deletion (Figure 15B). This deletion, presumably generated by the fast neutron mutagenesis, was located within a single gene model, Glyma.20G019300. The nearest ortholog to this gene in Arabidopsis is *NAPI*, or *Nck-Associated Protein 1* (gene model AT2G35110), and is involved in the actin cytoskeleton formation (Deeks et al., 2004; El-Assal et al., 2004). The soybean Glyma.20G019300 gene has high amino acid similarity to this Arabidopsis *NAPI* ortholog (87.4%). Furthermore, the Arabidopsis *nap1* trichome mutant (*grl-4*) has a phenotype with swollen, twisted, and shorter trichomes, similar to the soybean mutant phenotype observed in R55C01 (Supplemental Fig. 3B).

To validate the mapping results, co-segregation between the trichome phenotype and the candidate mutation was tested on segregating F₃ progeny. A co-dominant PCR marker was designed using three primers to amplify unique bands for the mutant and wild-type alleles (Table 7, Figure 13A). The phenotypes of 50 F₃ individuals, representing different F_{2:3} families, perfectly co-segregated with their expected genotypic classes (mutant: mt/mt or wild-type: wt/wt or wt/mt) (Figure 13B). Altogether, these data and prior information about the *NAPI* gene indicated that Glyma.20G019300 was the leading candidate gene for the soybean *gnarled* phenotype.

Complex rearrangements detected in Glyma.20G019300

The aCGH data detected an approximately 2 kb deletion in the Glyma.20G019300 candidate gene, but was not able to resolve the fine structure of this event. Therefore,

whole genome resequencing was conducted to resolve the specific breakpoints of this deletion. Surprisingly, the resequencing data revealed a much more complex structure to this locus than anticipated (Figure 16).

Glyma.20G019300 consists of 23 exons stretched across 22,550 nucleotides (including exons, introns, and untranslated regions) that encode a protein with 1,388 amino acids. Sequencing of the mutated allele resolved two distinct deletions (2,021 bp and 1,224 bp, respectively) nearby one another. The 1,224 bp deletion was not originally detected by aCGH due to the deletion of only a single probe, but occurred upstream of the aCGH identified mutation. Secondly, the 1,138 bp segment separating these deletions was found in an inverted orientation. This inverted fragment consisted of the thirteenth exon and part of the twelfth and thirteenth introns. This fragment inverted and fused to a sequence 7 bp upstream of the fourteenth exon, forming a novel junction between positions Gm20:2,010,290 and Gm20:2,007,928. Lastly, an unresolved chromosome rearrangement, possibly an intra-chromosomal translocation, was identified adjacent to the inversion-deletion junction. The Gm20:2,009,152 side of the inverted fragment was connected to sequence found nearly 15 Mb downstream in the reference genome, at Gm20:16,920,485. The first half of the gene was interrupted at Gm20:2,012,311 and was found to be connected to Gm20:16,939,673. A novel 22 bp sequence was also found in the junction. PCR primers were designed to span these junctions, and the amplicons confirmed the presence of the three novel junctions found by whole genome resequencing (Figure 16).

The putative intra-chromosomal translocation remains unresolved. Orientation of the sequences at the junction Gm20:2,009,152 to Gm20:16,920,485 and the junction Gm20:2,012,311 to Gm20:16,939,673 do not suggest that a single contiguous piece was inserted into Glyma.20G019300. Rather, the orientations of the sequences suggest that additional chromosome rearrangements have likely occurred.

RNA sequencing expression analysis of the *gnarled* mutant

Complex rearrangements may alter transcription of the genes at and nearby the disrupted locus. To test this, RNA-seq analysis was conducted on leaf, seed, and root tissues from wild-type ('M92-220') and *gnarled* mutant individuals (R55C01). Approximately 95% of the RNA-Seq reads were mapped to the soybean genome, and roughly 89% of the read-pairs were mapped concordantly. Under normal conditions in wild-type plants, Glyma.20G019300 (Glyma20g02370 in previous genome releases) has been observed to be transcribed in all previously examined tissues (Severin et al., 2010; Libault et al., 2010). Similarly, this gene was observed to be transcribed in all three tissues tested in this study, however the *gnarled* mutant exhibited an overall decrease in transcript abundance in the mutant individuals compared to wild-type (Figure 17).

Exons one through ten all exhibited similar transcription levels in mutant and wild-type plants, with some minor exon-specific fluctuations among the different tissues and genotypes, while exon 11 appeared to show increased relative transcription levels in the mutant (Figure 17). Exons 12, 13 and 14, however, exhibited dramatic differences

between the mutant and wild-type individuals (Figure 17). In each of these exons, the *gnarled* mutant exhibited essentially no transcription, while the wild-type individuals exhibited transcription levels similar to the other exons of this gene. As described in the previous section, exon 12 resides within a deleted interval in the *gnarled* mutant, and an interval encompassing exon 13 is inverted relative to wild-type. These structural rearrangements may be expected to eliminate transcription in these intervals. Exon 14, while present and in proper orientation in the mutant, is directly adjacent to the exon 13 inversion and also appeared to be transcriptionally silent. The downstream exons, 15-23, appeared to be generally downregulated in the *gnarled* mutant relative to wild-type (Figure 17). Furthermore, it is worth noting that the transposition event that separates exons 11 and 13 in the *gnarled* mutant is of unknown size and sequence composition (Figure 16). Therefore, it is possible that the reads observed from exons 1-11 and exons 15-23 are completely independent transcripts in the *gnarled* mutant line. Furthermore, gene model Glyma.07G221000, the nearest paralog to Glyma.20G019300, did not exhibit any exon-level transcript differences between the mutant and wild-type plants. This offers additional evidence that the transcriptional changes observed in Glyma.20G019300 are a consequence of the structural rearrangement *per se*, rather than RNA-interference or other post-transcriptional silencing mechanism.

Transcriptional alterations were also observed genome-wide between the *gnarled* mutant compared to wild-type, including 354 genes downregulated and 312 upregulated in leaves. These genes were involved in processes such as lipoxygenase activity (GO:0016165) and photosynthetic processes (GO:0009769 and GO:0016168), which are

pathways that have been linked to trichome development (Schillmiller et al., 2010; Yan et al., 2012). However, homologs for genes previously demonstrated to be involved in trichome developmental processes, including monosaccharide and lignin biosynthesis (Marks et al., 2009), were not observed to be differentially transcribed between the *gnarled* mutant and wild-type in any tissue (data not shown).

Complementation of the *Atnap1* using *GmNAPI*

Due to the high similarity of both the *nap1* gene sequences and phenotypes, complementation of an Arabidopsis *nap1* mutant with *GmNAPI* could be used to validate the function of Glyma.20G019300. A construct consisting of 2 kb of the soybean *NAPI* promoter driving the soybean *NAPI* cDNA and a D35S promoter driving the BAR herbicide resistance gene was transformed into the Arabidopsis *nap1 gnarled* mutant (*gri-4*). Twenty T₁ individuals were recovered that displayed wild-type trichomes, were resistant to Glufosinate, and tested positive for the *GmNAPI* transgene based on PCR analyses (Figure 18). The functional complementation of the Arabidopsis *nap1* mutant indicates that *GmNAPI* is important for trichome formation and the two orthologs share functional homology.

Identification of a spontaneous *NAPI* soybean mutant T31 (*p2*)

A search for historic soybean mutants with *gnarled* trichomes led to the identification of the mutant line T31 (PI548159) (Stewart and Wentz, 1926; Bernard and Singh, 1969; Singh et al., 1971; Healy et al., 2005). T31's recessive *p2* mutant trichome allele was initially described as 'puberulent' but more closely resembles the *gnarled* phenotype. The *p2* allele was previously backcrossed into the cv. 'Harosoy' (PI548573) and into cv. 'Clark' (PI548533) to generate several advanced backcross lines (Weiss and Stevenson, 1955; Johnson, 1958; Bernard et al., 1991). There are two *p2* backcross (BC₆) near-isogenic lines (PI547713 and PI547743) in 'Harosoy', and there are three *p2* backcross (one BC₆ and two BC₇) near-isogenic lines (PI547449, PI547565, and PI547566) in 'Clark' (Bernard et al., 1991). The SoySNP50K chip data (Song et al., 2016) obtained from SoyBase (<http://soybase.org>) was used to identify a genomic interval shared by T31 and the five *p2* backcross lines.

A single genomic interval shared between T31 and the five *p2* backcross lines was located on chromosome 20 (Figure 19). At position Gm20:1,742,275 (ss715636805), all five *p2* backcross lines carried the T31 allele, and at position Gm20:2,053,056 (ss715636914) three of the five *p2* backcross lines contained the T31 allele and two lines (PI547449 and PI547565) had genotyping fails. PI547565 had the T31 allele for the polymorphic SNP at Gm20:2,148,735 (ss71563945), adjacent to the failed SNP at Gm20:2,053,056, suggesting that the genotyping score at Gm20:2,053,056 would likely match T31. PI547449 had either missing data or heterozygous calls at all polymorphic

SNP positions downstream of position Gm20:1,742,275 until Gm20:2,353,994 which had the ‘Clark’ allele. The observed heterozygous genotype calls are likely due to heterogeneity found between sampled individuals in a line rather than to residual heterozygosity within a specific individual of a line. The narrow 566 kb *p2* introgression interval, marked by the resumption of the recurrent parent haplotypes upstream at Gm20:1,582,950 (ss715636740) and downstream at Gm20:2,148,735 (ss715636945) contains the *GmNAPI* gene Glyma.20G019300. The inclusion of Glyma.20G019300 in the *p2* introgression interval and the similarity of the R55C01 and T31 trichome phenotypes suggested that *p2* could be caused by a mutation in Glyma.20G019300.

Sequencing the exons of Glyma.20G019300 from T31 identified a single base pair deletion in the 22nd exon (Figure 20). The resulting frame shift mutation and early stop codon resulted in the mutation or loss of 202 amino acids (14.5% of the gene). Sequence analysis of the locus in 25 wild-type diverse soybean accessions (McHale et al., 2012) confirmed that the single base pair deletion is unique to T31.

DISCUSSION

The morphological development of trichomes is guided by the actin cytoskeleton and thus proper actin nucleation is critical for proper trichome formation (Beilstein and Szymanski, 2004). The *NAPI* gene was identified as a component of the SCAR/WAVE protein complex that activates the ARP2/3 complex involved in actin nucleation, and

mutations in this gene have been shown to result in the *gnarled* trichome phenotype (Deeks et al., 2004; El-Assal et al., 2004).

Several previous studies have described soybean trichome morphological mutants (Nagai and Saito, 1923; Piper and Morse, 1923; Stewart and Wentz, 1926; Owen, 1927; Johnson and Hollowell, 1935; Ting, 1946; Williams, 1950; Bernard and Singh, 1969; Bernard, 1975; Healy et al., 2005). However, to our knowledge, no soybean study has cloned the underlying causative variant of a soybean trichome morphology mutant. This study combined aCGH and WGS-BSA to identify that fast neutron induced chromosome rearrangements in Glyma.20G019300, the soybean *NAPI* ortholog, caused the R55C01 mutant's *gnarled* trichome phenotype. The functional complementation of an *Arabidopsis nap1 gnarled* mutant (*grl-4*) by whole plant transformation with *GmNAPI* validated the soybean *GmNAPI* gene function. Additionally, this study further validated the function of *GmNAPI* by identifying a second mutant allele of *GmNAPI*, the *p2* trichome mutant locus in line T31. This phenotype is the result of a single base pair deletion in the 22nd exon of *NAPI* (Figure 20).

This study's 26.6 Mb mapping interval using WGS-BSA was wider than expected, but the size of the interval was likely inflated by low regional recombination rates. The mapping interval had reasonable resolution on the distal side of the candidate gene with the mapping interval starting approximately 329 kb from *GmNAPI*. However, on the proximal side of *GmNAPI*, the mapping interval extended approximately 26.3 Mb from *GmNAPI* to the other arm of the chromosome. Due to the position of *GmNAPI* near the edge of the heterochromatic region, it was likely that repressed recombination on the

proximal side of the gene expanded the mapping interval significantly. A recent mapping study also identified suppressed recombination in this region on chromosome 20 (Li et al., 2014). We suspect that the repressed recombination in this region led to the large mapping interval, and it is likely that subsequent mapping studies using WGS-BSA in soybean will have smaller mapping intervals for regions of the genome with higher recombination rates.

Combining the aCGH data with the WGS-BSA mapping interval led to the identification of a single candidate gene, despite the large mapping interval. The only aCGH-detected mutation in the mapping interval was the approximately 2 kb deletion in Glyma.20G019300. Further examination at this locus identified additional mutations that were not detected by aCGH. These additional mutations include a second deletion, an inversion, and two novel junctions which suggest additional chromosome rearrangements occurred on chromosome 20. The complexity of fast neutron induced mutations identified within this single gene was unexpected and further challenges the common assumption that fast neutron mutagenesis results in simple deletions (see Bolon et al., 2014 for additional evidence).

The R55C01 mapping results were validated using a combination of whole plant transformation and identification of a second mutant allele. Functional complementation of an *Arabidopsis nap1* mutant with *GmNAPI* validated the candidate gene. The identification of a second soybean allele, a frame shift mutation in *GmNAPI* in the *gnarled* trichome soybean mutant T31 (PI548159), further validated the candidate gene hypothesis.

This study has demonstrated the effective combination of WGS-BSA and aCGH to identify a candidate fast neutron induced mutation from a reasonable sized F₂ mapping population. The combination of technologies demonstrated the ability to save significant cost and time by identifying the causative variant with only one round of BSA mapping, and without the need for additional fine-mapping.

CHAPTER 4

Map-based cloning and physiological characterization of the soybean *lps1* short petiole locus

SUMMARY

Increases in the world population have led to rising demands for sustainable sources of plant protein and oil. Modifications to plant canopy architecture could play an important role in meeting these demands, as was demonstrated during the green revolution in the 1960's and 1970's. Petioles are an important component of canopy architecture as they connect the leaflets to the stem in dicot species, but many of the mechanisms controlling petiole length are largely unknown. To our knowledge, no study has conducted a thorough yield and physiological study on the effect of short petioles on soybean yield. Kilen (1983) identified a short petiole soybean mutant, *lps1*, that segregated as a single, recessive locus. The mutation was first observed in 1976 segregating in an F₃ row in a population crossed between Forrest(2) x (PI 229358) and D71-6234. D71-6234 was derived from a cross between a high protein Lee type and PI95960. None of the parents were observed to have the short petiole phenotype. This study was undertaken to identify the causal DNA polymorphism underlying the *lps1* mutation as well as to assess the physiological and agronomic effects the *lps1* mutation using Near-Isogenic Lines (NILs) generated from heterogeneous inbred families during the trait mapping process. Bulk segregant analysis with whole-genome

re-sequencing data located the chromosomal region underlying this trait. Diversity analysis of genes within this interval has identified a candidate mutation that may cause this unique architecture. Preliminary yield trials indicate that in narrow row spacings, the *lps1* NILs on average displayed an increase in both a grain yield and harvest index. These preliminary results suggest that the *lps1* phenotype could be utilized to improve soybean yields when cultivated in narrow row spacings.

INTRODUCTION

The architecture of the plant canopy can greatly affect the productivity a crop production system; and through the centuries, farmers and plant breeders have increase crop productivity by selecting plants with modified canopy architecture. Breeding efforts that modify a plant's canopy architecture by changing how a plant allocates its energy fundamentally affects harvest index (Long et al., 2006), and increases in grain yield have historically been accompanied by increases in harvest index (Evans, 1997; Grifford et al., 1984; Hay, 1995). In wheat, examples of changes to canopy architecture include decreased plant height, increase flag leaf area, and increased harvest index (Jiang et al., 2003). A survey of 100 years of barley improvement identified that the newer cultivars had higher harvest index (Riggs et al., 1981). Increases in soybean grain yield have also been correlated with increases in harvest index. A survey of physiological changes that occurred over 58 years of short season soybeans cultivar releases found a linear 0.47% per year increase in harvest index with year of release of the cultivars (Morrison et al.,

1999). Suhre et al. (2014) conducted a similar experiment surveying 80 years of soybean cultivar releases in Maturity Groups (MGs) II and III and also found a correlation between increased harvest index and increased grain yield. In contrast to wheat, barley and soybean, it has been suggested that maize harvest index has not increased as yields have increased, but rather that an increase in biomass has led to increased maize yields (Lorenz et al., 2010). The correlated increase in harvest index with increased grain yield across many species suggests that modifying canopy architecture to achieve increased harvest index is a viable method for increasing crop yields.

It may be possible that yield improvements in soybean could be produced through modifications to canopy architecture such as decreasing the length of the petiole, the structural organ that connects leaves to the stem or branch. In a modeling study with tomato, Sarlikioti et al. (2011) found that increasing petiole length diminished both canopy photosynthesis and light absorption.

The soybean short phenotype (*lps1*) (Figure 21A) was first observed in 1976 in three of 422 F₃ lines from a population derived from a cross between Forrest(2) x (PI 229358) and D71-6234 (Hartwig and Epps, 1973; Kilen, 1983). D71-6234 was derived from a cross between a high protein Lee type and PI95960 (Johnson, 1958). In 1976, a single short petiole individual was chosen from the F₃ row number 1609 and advanced by single seed descent to the F₆ generation. The single F₆ short petiole row was bulked and named D76-1609 (T279, PI 548256). None of the four parents were observed to have the short petiole phenotype suggesting that the *lps1* mutation arose spontaneously. Kilen

(1983) found that *lps1* phenotype segregated as a single, recessive nuclear locus and also speculated that *lps1* could be a desirable trait for soybean improvement.

The development of dense genotype platforms and the resequencing of whole genomes has allowed the use of new mapping methodologies to identify and clone desirable traits. For instance, the use of Bulk Segregant Analysis (BSA) (Michelmore et al., 1991) using genetic sequence (RNA, whole-genome or exomes) has been demonstrated in a number of plant species including tomato (Illa-Berenguer et al., 2015), maize (Liu et al., 2012; Haase et al., 2015), rice (Takagi et al., 2013; Yang et al., 2013), barley (Mascher et al., 2014) and Arabidopsis (James et al., 2013; Zhang et al., 2014).

The objectives of this study were to identify the *lps1* causative mutation and to characterize the physiological and agronomic effects of this mutation. Additionally this study wanted to test the ability to conduct whole genome sequenced based bulked segregant analysis using bulking across populations.

MATERIALS AND METHODS

Mapping population development

The *lps1* mutant line D76-1609 was outcrossed to five genetically diverse wild-type lines adapted to MG 0 and I: Heinong51, Kengfeng16, MN1410, Norchief, and Ozzie (Johnson, 1958, Orf et al., 1985). F₁ plants were grown in the greenhouse, and a large number of F₂ individuals were grown in a Chile winter nursery but were not phenotyped. Due to the drastic differences in maturity of the parental lines, the D76-

1609 mutant line is a maturity group MG VII soybean and the five wild-type lines are maturity group 0 and I maturities, selection was done for early maturing F₂ individuals. F_{2:3} families were planted at St. Paul and visually phenotyped for petiole length. Families were scored as homozygous mutant, segregating, or homozygous wild-type.

Whole Genome Sequencing based Bulked Segregant Analysis

Whole Genome Sequencing based Bulked Segregant Analysis (WGS-BSA) with bulks of 50 plants each was used to map the *lps1* mutation. Two bulks were constructed, mutant and wild-type, with each bulk containing an equal representation of individuals from the five mapping populations. The bulks were generated by collecting tissue from single individuals from rows that were homozygous mutant or homozygous wild-type. Single leaf punches were collected from each leaf, and DNA was extracted from the bulked leaf punch sample using a Qiagen DNeasy kit. Bulks were sequenced to an average sequence depth of 30x at the UMGC. Reads were aligned to Glyma.Wm82.a2.v1 (Song et al., 2016), and allele frequencies were called for the SoySNP50K positions (Song et al., 2013).

Genotyping of the bulks was conducted by the same method described in the 3rd chapter of this thesis. Before mapping, non-polymorphic SNPs (having allele frequencies less than or equal to 0.1 or greater than or equal to 0.9 in both mutant and wild-type bulks), SNPs with missing data in either bulk, and SNPs with read counts less than ten in either bulk were removed from the dataset. Next, the allele frequencies were plotted and analyzed to identify a spread in allele frequency between the mutant and wild-type bulks.

In this case, the spread in allele frequency is caused by the mutant bulk moving to fixation around the *lps1* causative locus. Due to the complexity generated by bulking across five populations, the absolute differences between the mutant and wild-type bulk reference allele frequencies was used to identify a chromosome region a large spread in allele frequency. An allele frequency absolute difference of 0.66 was used to determine the edges of the mapping interval.

Population advancement From F_{2:3} to F_{3:4}

Wild-type individuals from segregating F_{2:3} families that matured early enough to be adapted to MG II and earlier, were primarily selected to advance to the F_{3:4} generation. Advancing primarily wild-type individuals from segregating rows was done to enrich the F_{3:4} population for individuals segregating for *lps1*. Selection for maturity was done to increase the percentage of the population that was adapted to the northern maturity zone. After selection for maturity, 320 families were advanced.

Fine mapping of the *lps1* locus.

The 320 F_{3:4} families were planted and phenotyped for petiole length. From these families 1,473 individual plants representing all five populations were tagged and leaf tissue was collected. DNA was extracted using a Qiagen BioSprint 96 DNA Plant Kit following the manufacturer's protocol. The F_{3:4} individuals and the six parents were genotyped at the University of Minnesota Genomics Center (UMGC) core facility using a

custom panel of 58 SNP assays on a Sequenom MassARRAY genotyping platform that were positioned across the BSA mapping interval. The phenotype and genotype data were then compared and analyzed to identify the interval containing the *lps1* locus.

Identification of the candidate polymorphism

Once the fine mapping interval had been identified, the sequences of the bulks were compared against the Glyma.Wm82.a2.v1 (Song et al., 2016) sequence and polymorphisms were identified in the mapping interval. In addition, the allele frequencies of these polymorphisms were calculated. The polymorphisms were compared against the polymorphisms present in seven wild-type lines previously resequenced at the UMGC or had sequence data available online at NCBI ('Archer', 'Noir 1', 'Minsoy', 'Williams', IA3023', M92-220, and *G. soja* var. IT182932) (Bernard and Lindahl, 1972; Cianzio et al., 1991; Orf and Denny, 2004; Kim et al., 2010), to identify polymorphisms that were (1) present in the mutant bulk and not present in the wild-type lines and (2) occurred at a high frequency in the mutant bulk and low frequency in the wild-type bulk.

As the D76-1609 *lps1* phenotype occurred spontaneously, the eight candidate polymorphisms were screened on D76-1609 and its four parents (Forrest, PI229358, Lee, and PI95960) to identify a polymorphism found in D76-1609 that was not found in any of the parents. DNA was extracted from all lines using a Qiagen DNeasy kit. Cleaved amplified polymorphic sequences (CAPS) assays were used to screen the SNPs and one small insertion polymorphism (Konieczny and Ausubel, 1993). The deletion

polymorphisms were genotyped by simple PCR amplification followed by band size separation on an agarose gel. The primers and restriction enzymes used for these reactions are provided (Table 11). A single insertion (T/TAAG) in Glyma.16G209100 was then screened on a diverse panel of 29 wild-type soybean lines (McHale et al., 2012) to see if the polymorphism was found in other soybean lines. A synonymous polymorphism identified in Glyma.16G209100 was also screened on 27 diverse soybean lines (McHale et al., 2012) to identify the frequency of that polymorphism in soybean germplasm.

Segregation test of the *lpsI* mutant with mutant phenotype

To test the cosegregation of candidate mutation with plant phenotype, F_{4:5} individuals were selected to be phenotyped and genotyped specifically for the AAG insertion. F_{3:4} genotype and phenotype data was used to select a panel of individuals representing all five populations and were composed of (1) individuals that were predicted to have recombinations in the fine mapped interval, or (2) individuals for which the F_{3:4} genotype did not match F_{3:4} phenotype. The families were all phenotyped, and the six parental lines and 88 individuals representing unique F_{4:5} families were genotyped for the AAG insertion using the CAPS assay as previously described.

Selection and advancement of F_{3:4} individuals to development F_{4:6} NILs for preliminary yield trials

Selection was done to identify Heterogeneous Inbred Families (HIFs) (Tuinstra et al., 1997) from which to generate NILs. F_{3:4} families identified as segregating for petiole length were also visually inspected to select families uniform in agronomic characteristics such as height and maturity. The same genotyping data used to fine map the *lps1* locus was additionally used to identify homozygous wild-type individuals and confirm the visually identified homozygous recessive *lps1* individuals from the selected HIFs. Twelve pairs of F_{4:5} mutant and wild-type individuals were advanced to 8m Chile winter nursery rows for seed increase.

F_{4:6} Preliminary yield trials

Due to the short petiole lines' slow rate of canopy closure on 76 cm wide rows, the yield trial plots were ten, 25 cm rows wide by 3.6m long. The plots of each *Lps1/lps1* NIL pair were planted adjacent to one another in order to minimize environmental variation between the pair of lines, and the location of the paired plots were randomized within locations. The two locations that were planted were St. Paul and Rosemount. Some NIL pairs had two replications per location while the majority of NILs had only a single replication per location due to limitations in the seed amount produced at the winter nursery. Seeding rates were set at approximately 500,000 seeds per hectare. Height and growth stage notes were recorded during the season at the St. Paul location only, and lodging notes were recorded before harvest at both locations. A bordered

0.76m² sample area 1m long and 3 rows wide was hand harvested by cutting the plant stems at ground level. The total mass of the plants was weighed before threshing in order to be able to calculate harvest index. After all of the plot samples were harvested, each plot was harvested with a small plot combine in order to collect seed for future yield trials. Harvest index was calculated as the mass of the cleaned, threshed seed divided by the total plant biomass harvested from the 0.76 m² sample area. 100-seed weights were measured by weighing five sets of 100 seeds. All measurements were calculated as the *lps1* mutant NIL relative to its *Lps1* wild-type NIL. Statistical comparisons were done using paired t-tests. Some pairs of NILs were dropped from the Rosemount location analysis due to one or both of the pair of plots being damaged by pocket gophers.

Advancement of plant rows from F_{3:4} to F_{5:6}

The same genotyping data used to fine map the *lps1* locus was additionally used to identify individuals that were predicted to be heterozygous in the *lps1* mapped interval. Visual selection was also done to advance lines with more favorable agronomics such as early maturity, low lodging, and high breeder score. 81 individuals with good agronomic characteristics and were genotyped as heterozygous for part or all of the *lps1* mapping interval were advanced to Chile winter nursery to be grown on individual 2m rows. Of these 81 individuals, eighteen were also planted into a second row for a total of 99 rows. From each F_{4:5} row in Chile, up to three short petiole plants were tagged during the season and then at maturity, all tagged plants were harvested plus seven to ten wild-type plants to reach a total of ten F_{5:6} plants harvested per row.

F_{5:6} Near Isogenic Lines

The following summer, the F_{5:6} lines were planted into individual 3m rows and phenotyped for petiole length, flower color, pubescence, and leaf shape in order to identify any lines with contaminants. The field was also walked to identify pairs of lines descending from the same F_{4:5} heterozygous plant that had similar agronomic characteristics such as height and maturity but differed for petiole length. These individual pairs were used for physiological characterization of the *lps1* mutant phenotype.

Petiole cell lengths were measured to determine if changes in cell lengths had occurred. Petiole lengths were measured on four separate *Lps1/lps1* NIL pairs using the petioles on the 6th and 7th fully opened leaves from the top of the plant. Petioles were measured from three plants per NIL. Petioles were cut into three pieces: distal, middle, and proximal and cleared in 1M KOH at room temperature. When tissue was adequately cleared, the middle piece from each plant line was rinsed three times in distilled water. To prepare for cell size evaluation, portions of the epidermal peel were removed from the center portion of the middle piece and stained with safranin for two to five minutes. The stained tissue was rinsed three times with distilled water. Cell length values were obtained from five individual epidermal cells from three different locations along the stained epidermal peel. Measurements were made from three petioles from each line of the four separate *Lps1/lps1* NIL pairs for a total of 45 measurements per NIL and a total of 360 measurements. Images were obtained using the Nikon DS-Ri1 color camera

attached to the Nikon Eclipse NiU light microscope. All measurements were obtained using the NIS Elements AR software.

RESULTS

Gross Mapping of *lps1* mutant to Chromosome 16

To map the *lps1* causative variant, the *lps1* mutant line D76-1609 was outcrossed to five genetically diverse wild-type lines: Heinong51, Kengfeng16, MN1410, Norchief, and Ozzie. F₁ individuals displayed a wild-type petiole phenotype, indicating that the short petiole phenotype is recessive (data not shown). F_{2,3} families were planted in St. Paul and visually phenotyped for petiole length. While individual heterozygous F_{2,3} families were found to segregate 1:2:1 (homozygous wild-type : heterozygous : homozygous short) (data not shown) as expected for a one locus trait as was found by (Kilen, 1983), across the populations the F_{2,3} families did not segregate 1:2:1. Instead, across the populations *lps1* segregated 266 : 234 : 38 (homozygous wild-type : heterozygous : homozygous short) (Chi-squared $p = <0.001$). This result could be due to several factors. First, a main reason could be that the selection done in the winter nursery to advance early maturing individuals may have led to a decrease in the number of individuals with the mutant phenotype, as D76-1609 is a maturity group VII line. Second, the homozygous *lps1* individuals with shorter petioles are not able to compete as well with long petiole plants within a row. If that were the case, the homozygous *lps1* F₂ individuals may not have been advanced from Chile. This same factor, could have

decreased the survival of homozygous *lps1* F_{2:3} individuals in heterozygous families until the petiole length phenotyping which occurred late in the season when the petiole length differences become clear. Not advancing homozygous *lps1* plants from Chile would have resulted in a decrease in the number of homozygous *lps1* F_{2:3} families. The loss of homozygous *lps1* individuals from heterozygous F_{2:3} families would cause an increase in the number of homozygous wild-type rows and a resulting decrease in the number of heterozygous rows.

We were intrigued to test whether bulking across populations using Whole Genome Sequencing based Bulk Segregant Analysis (WGS-BSA) could be used to effectively map the *lps1* mutation. In the mutant and wild-type bulk samples, it is expected that the mutant and wild-type parent genotypes will be present in approximately equal proportions for regions of the genome not associated with the causative variant. In contrast, a chromosomal region containing the causative variant is not expected to have an equal proportions of the mutant and wild-type parent genotypes in the bulks. Thus the causative locus can be identified by large spreads in allele frequency between the mutant and wild-type bulks. The absolute differences between the reference allele frequencies of the bulks was used to identify chromosome regions with spreads in allele frequency (Figure 22). A single genomic region with a large spread in allele frequency was observed on the bottom end of chromosome 16. An absolute difference of 0.66 in allele frequency was used to determine the edges of the mapping interval. The mapping interval identified was (35.8Mb to 37.5Mb) or a 1.7Mb interval on chromosome 16.

Fine Mapping of the *lps1* mutation

While some homozygous mutant rows were advanced to the next generation, primarily wild-type individuals from segregating F_{2:3} families that matured early enough, were selected to advance to the F_{3:4} generation. From these segregating rows, it is expected that 66% of the long petiole individuals were heterozygous for *lps1*. Thus, by primarily advancing wild-type individuals from segregating rows, the F_{3:4} population would be enriched for families segregating for *lps1*. Using this method, 43% of the F_{3:4} families (138/320) segregating for *lps1* instead of 25% of rows segregating as would be expected if the population were advanced by single seed descent.

Two custom Sequenom assays totaling 58 SNPs spread across the BSA mapping interval were screened on 1,473 F_{3:4} individuals and the six parental lines to fine map the *lps1* locus. This fine mapping identified a 84.9 kb interval containing 11 genes (Figure 23A-B).

Candidate gene identification

The sequences of the bulks were compared against the Glyma.Wm82.a2.v1 sequence (Song et al., 2016) to identify 950 polymorphisms in the 84.9kb mapping interval. These polymorphisms were screened to identify polymorphisms that were (1) present in the mutant bulk and not present in a panel of seven previously resequenced wild-type lines and (2) were present at high frequency in the mutant bulk and low frequency in the wild-type bulk. From this analysis eight polymorphism were identified:

three SNPs and five INDELS (Table 11), some of which were predicted to be synonymous changes or were intergenic changes.

The *lps1* phenotype in D76-1609 appeared to have arisen spontaneously (Kilen, 1983), and thus the causative variant is not expected to be present in the parents of D76-1609. Haplotype analysis of SoySNP50K data (Song et al., 2015) indicated that the fine-mapped interval was likely inherited from Forrest (Table 12). However, to be thorough, all 8 polymorphisms were tested against the four parents of D76-1609: Forrest, PI229358, Lee, and PI95960. Seven of the eight polymorphisms were found in at least one parent of D76-1609, and only one polymorphism was found to be unique to D76-1609 (Figure 23B). This unique polymorphism was a 3 bp in-frame insertion of an AAG (T/TAAG at Gm16: 36849837) into the second exon of Glyma.16G209100 (Figure 23C). This change adds an extra lysine at amino acid position 480. In addition to the 4 diverse parents and Williams 82, an analysis of 29 diverse long petiole soybean lines found that the AAG insertion was unique to D76-1609 (Figure 24). As a comparison, a synonymous SNP (C/T at Gm16:36848410) 1,428 bp upstream of the AAG insertion was identified in the 1st exon of Glyma.16G209100 and was tested on the diversity panel. This polymorphism was found in D76-1609, Forrest, PI229358 and 6 of the 27 (22%) diversity panel lines tested (data not shown) indicating that the variant is not rare in the diversity panel.

Segregation test of the *lps1* mutant with mutant phenotype

In order to test the cosegregation of the mutant phenotype with the genotype, F_{4.5} individuals were selected to be phenotyped and genotyped specifically for the (T/TAAG)

insertion. Individuals were selected using the F_{3:4} genotype data. Specifically, the selected panel of individuals representing all five populations was composed of individuals that (1) were predicted to have recombinations in the fine mapped interval or (2) for which the F_{3:4} genotype did not match F_{3:4} phenotype. The families were phenotyped and 88 individuals representing unique F_{4:5} families and the six parental lines were genotyped. For seventeen lines (19.3%), the 2014 phenotype data disagreed with the 2014 genotype data, but agreed with the 2015 phenotype and genotype data. For three of these lines (3.4%), the 2015 *Lps1* wild-type phenotype perfectly matched the 2015 *Lps1* 'T' genotype. It is suggested that these three lines were incorrectly tissue sampled in the 2014 field. Fourteen lines (15.9%) in 2014 had the *lps1* haplotype through the entire fine-mapped interval and also were homozygous for the *lps1* AAG insertion. These lines however did not strongly exhibit the mutant phenotype. In some cases, only a single petiole on the plant was observed to have the *lps1* short petiole phenotype (Figure 25), indicating that there may be an interaction of certain genotypes and environmental conditions. All other lines (81.0%) had perfect matches between the 2014 phenotype, 2015 phenotype, 2014 genotype, and 2015 genotype. These results validate that the *lps1* phenotype is caused by the identified AAG insertion but also indicate that environmental conditions may affect the expressivity of the *lps1* phenotype.

Candidate gene characterization

Glyma.16G209100 is listed as 'Plant protein of unknown function (DUF247)', and thus information about the structure, expression, and mutant analysis were gathered

in order to try to characterize the gene. Wild-type Glyma.16G209100 has 2 exons with a cDNA of 1,590 nucleotides which codes for 529 amino acids. The allele present in D76-1609 has an extra lysine at the 480 amino acid position for a total of 530 amino acids. The wild-type gene has a 103 bp 5'UTR ending directly upstream of the start site, and a 163 bp 3'URT is starting directly after the stop site. Two transmembrane domains were predicted in the wild-type allele from amino acid positions 126 to 135 and from 499 to 512 using the Dense Alignment Surface (DAS) algorithm (Cserzo et al., 2002). The DAS algorithm did not predict that the insertion of lysine disrupted these transmembrane domains other than to position the second transmembrane domain to start and end one amino acid later as the insertion occurs between the predicted transmembrane domains. Expression analysis indicates that Glyma.16G209100 is expressed in the leaf, flower, root, shoot apical meristem, stem, pod, and the mixed sample while showing low to no expression in the root hairs, nodules, and seed (Figure 26A) (expression data available on Soybase.org provided by Schmutz et al. [pers comm]). In contrast, expression data indicates that Glyma.09G159900, the paralog of Glyma.16G209100, shows low to no expression in all tissues (Figure 26B) (expression data available on Soybase.org provided by Schmutz et al. [pers comm]). Amino acid sequence analysis by TargetP, which predicts the subcellular cellular localization of a protein by amino acid sequence, suggested that the protein is not located in the chloroplast, mitochondrion, or in the secretory pathway with a reliability class of 3 (on a scale of 1 to 5, where 1 indicates the strongest prediction) (Emanuelsson et al., 2007). The closest ortholog in Arabidopsis, at 51% amino acid similarity is AT3G02645. T-

DNA mutants for AT3G02645 were obtained and planted out with wild-type controls to see if mutants in the ortholog would exhibit a short petiole phenotype; however, no differences in petiole length were observed. Due to the low similarity between the orthologs, it is not necessarily surprising that the Arabidopsis mutants did not display a short petiole phenotype.

Agronomic comparisons of *Lps1* and *lps1* NIL pairs

The *Lps1/lps1* F_{4:6} NILs were compared for several agronomic traits such as height, growth stage, and lodging. During the year, plant height measurements were taken at the St. Paul location each week from the end of June to the beginning of August (Table 13). These results indicate that for most of the vegetative growth period, there was no significant difference in NIL height, but starting in the V7 and continuing into the reproductive growth stages, the short petiole NILs were statistically significantly shorter than the long petiole NILs (Table 13). Growth staging notes taken during the season indicated that the NIL pairs did not show statistically significant differences in their growth stage (Table 13). Across the pairs of NILs, the short petiole lines had a non-significant mean relative increase of 0.25% ($p = 0.1607$) in lodging score (Table 14). Thus for three agronomically important traits, the *Lps1/lps1* NILs only showed statistically significant differences for relative height, and specifically relative height during the start of the reproductive stages.

Yield and quality comparisons

In addition to agronomic comparisons, the homozygous *Lps1* and *lps1* F_{4:6} NILs were compared for several seed quality and yield related traits (Figure 27; Table 14, Table 15, and Table 16). Across the NIL pairs, the *lps1* NILs had a non-significant mean relative increase of 1.0% ($p = 0.1157$) in oil and a non-significant relative mean decrease of 0.9% ($p = 0.1098$) in protein. Looking at performance by specific location compared to *Lps1* NILs, the *lps1* NILs at the St. Paul location had a non-significant relative mean decrease of 0.7% in seed protein ($p=0.3514$) and a non-significant relative mean increase of 1.0% in seed oil ($p=0.2677$) while at the Rosemount location, the *lps1* NILs had a non-significant relative mean decrease of 1.1% in seed protein ($p=0.1728$) and a non-significant relative mean increase of 1.0% in seed oil ($p=0.2829$). For grain yield, the *lps1* NILs had a non-significant relative mean increase of 8.9% ($p = 0.0696$). Looking at performance by specific location compared to *Lps1* NILs, the *lps1* NILs at the St. Paul location had a non-significant relative mean increase of 14.3% in grain yield ($p=0.0615$) while at the Rosemount location, the *lps1* NILs had a non-significant relative mean increase of 1.6% in grain yield ($p=0.7647$). For 100-seed weight, the *lps1* NILs had a statistically significant relative mean decrease of 4.8% ($p=0.0008$) in 100-seed weight. Looking at performance by specific location compared to *Lps1* NILs, the *lps1* NILs at the St. Paul location had a non-significant relative mean decrease of 2.9% in 100-seed weight ($p=0.1650$) while at the Rosemount location the *lps1* NILs had a significant relative mean decrease of 7.3% in 100-seed weight ($p<0.0001$). For harvest index, a measure of the

partitioning efficiency calculated as the aggregate biomass energy apportioned into the harvested fraction of the crop, the *lps1* NILs showed a significant relative mean increase of 10.2% ($p = 0.0146$). Looking at performance by specific location compared to *Lps1* NILs, the *lps1* NILs at the St. Paul location had a significant relative mean increase of 14.3% in harvest index ($p=0.0357$) while at the Rosemount location, the *lps1* NILs had a non-significant relative mean increase of 4.7% in harvest index ($p=0.2057$). The statistically significant relative mean increase in harvest index in the short petiole NILs compared to their long petiole NILs indicates that the short petiole trait is increasing the soybean plant's partitioning efficiency. For biomass production, as measured by the total mass of the seeds and plant harvested at the end of the season, the *lps1* NILs had a non-significant relative 0.1% increase in biomass accumulation ($p=0.7646$). Looking at performance by specific location compared to *Lps1* NILs, the *lps1* NILs at the St. Paul location had a non-significant relative mean increase of 2% in biomass accumulation advantage ($p=0.7646$) while at the Rosemount location, the *lps1* NILs had a non-significant relative mean decrease of 2.3% in biomass accumulation ($p=0.6157$). As most of the plants were harvested at the R8 growth stage after the petioles had fallen off, the higher harvest index for the *lps1* NILs indicates that morphological and physiological changes are occurring in the short petiole plants which led to a higher harvest index. The significant relative mean increase in harvest index by the *lps1* NILs was not accompanied by a significant relative mean decrease in their biomass accumulation suggesting that the *lps1* NILs are increasing harvest index by increasing the amount of seed produced per unit biomass rather than decreasing their biomass produced. These preliminary yield trial

results indicate that the *lps1* short petiole trait is not detrimental to grain yield and may in fact be beneficial through improving harvest index.

Physiological characterization of *lps1* mutants

Pairs of F_{5,6} NILs were observed in order to characterize the change in petiole length. Petiole length measurements from 4 separate *Lps1/lps1* NIL pairs indicate that the *lps1* mutation causes a relative mean decrease in petiole by 43.0% ($p < 0.0001$) (Figure 28A). Petiole cell lengths were also examined to determine if changes in cell length were contributing to petiole length differences. Across the four *Lps1/lps1* NIL pairs, the *lps1* NIL had statistically significantly decrease in cell length by 32.6% ($p < 0.0001$) (Figure 28B). This decrease in cell length is close to the average petiole length, suggesting that a decrease in cell length is an important factor in the mutants shorter petiole length.

During the microscopy work to measure cell lengths, it was observed that the cells in the *lps1* mutant petioles do not grow in an organized, linear fashion but exhibit a non-linear disorganized arrangement (Figure 21 B and C). This result could be caused by the mutant cells dividing in a combination of both transversely and abnormally orientated instead of the wild-type longitudinally and transversely. Additionally, the *lps1* mutant petioles exhibit a higher frequency of what appear to be collapsed or undeveloped cells as compared to wild-type petioles (Figure 21 B and C). It is possible that a combination of

shorter cells, non-linear disorganized arrangement, and the frequency of undeveloped or collapsed cells result in the shorter petiole length.

DISCUSSION

Through the centuries, farmers and plant breeders have selected plants with favorable canopy architecture modifications as a way to increase crop productivity. In soybean, some examples of canopy architecture modifications that have been selected and have recently been cloned include two different stem determinacy traits, *Dt1* and *Dt2*, as well as for *ln* trait with has pleiotropic effects on increased seed number per pod and lanceolate leaves (Liu et al., 2010; Tian et al., 2010; Jeong et al., 2012; Ping et al., 2014). Several petiole and petiolule mutants have been reported in soybean including: *lps1* (Kilen, 1983), *lps2* (You et al., 1998), *lps3* (Jun et al., 2009), a short petiole fast neutron mutant (Bolon et al., 2011), and a short petiolule mutant *lc* (Cary and Nickell, 1999), but to our knowledge, none of these petiole or petiolule mutants have been previously cloned. The main objectives of this study were to identify the causative variant underlying *lps1* short petiole trait in soybean and to physiologically characterize the mutant phenotype. Additionally, this study was used to test whether whole genome sequence based bulked segregant analysis (WGS-BSA) with bulking across populations could be used to map a qualitative trait. Here we report cloning of the *lps1* short petiole trait in soybean as due to an in-frame insertion of nucleotides AAG into the uncharacterized protein Glyma.16G209100, and we characterized some of the phenotypic

effects of the *lps1* mutation. The study also demonstrated that it is possible to use WGS-BSA with bulking across populations.

The identified in-frame AAG insertion was an unexpected casual variant for a recessive mutant which would suggest a loss of function mutation. The AAG insertion is positioned adjacent to a native AAG sequence (Figure 23C), and so it is possible that the AAG insertion occurred as a result of a DNA polymerase error or by an error created during a double strand break repair. At this time the function of Glyma.16G209100 is not clear nor is it clear how the in-frame AAG insertion disrupts gene function.

It is clear however, that the disruption of this gene affects cell elongation and development (Figure 21 and Figure 28). One potential explanation for the shorter petioles, shorter plants, and smaller seeds could be that this gene is involved in auxin sensing. Greer and Anderson (1965) found that shorter soybean petioles were produced on plants after applications of Triiodobenzoic acid (TIBA), an antiauxin that hinders auxin production. Additionally, they found that the application of TIBA caused a decrease in plant height, decreases in seed size, and in cases where yield was increased, they saw an increase in seed number. In high doses, TIBA decreased stem and petiole weights and there was an increase in the fraction of reproductive structures as a portion of the total plant dry weight. Put another way, high doses of TIBA increased harvest index. In this study, the observed shorter petiole cell lengths in the mutant NILs (Figure 28) additionally suggests that a disruption of auxin sensing could be giving rise to the shorter petioles, decreased plant height (Table 13), decrease in seed weight (Figure 27), and increase in harvest index (Figure 27). However, at this time the biochemical cause for

shorter petioles it is not known, and further testing is needed.

An objective of the study was to determine if WGS-BSA with bulking across populations could be used to map a trait. Such a strategy could be useful for mapping a trait that occurs at a low frequency in several populations. While this strategy was effective at mapping the trait, the bulking across populations was determined to add additional complexity to the mapping and it is not clear how this approach affected the mapping resolution. One downside of bulking across multiple populations was that in some places in the genome, one or more of the parental lines could share the allele of the D76-1609 and cause an artificially low difference of allele frequency. The occurrence of shared haplotypes was reduced by using northern germplasm as the outcross parents to map *lsp1* found in a southern germplasm line, because northern and southern germplasm are expected to be divergent for most haplotypes as they are not typically crossed to each other. The mapping approach could have been improved by first sequencing D76-1609 and the outcross lines to identify polymorphisms between the parents lines and then screening these polymorphisms on the bulks, rather than screening the SoySNP50K positions.

The disorganized cell pattern seen in the *lsp1* petiole epidermis cells could suggest that the mutation affects cell division. A mutant in maize, *tangled1*, exhibits a similar phenotype of disorganized cell arrangements (Smith et al., 1996; Cleary and Smith, 1998). Similar to the *tangled1* mutant, the *lps1* mutant appears to display normal leaf shape suggesting that the possible disruption in cell division does not affect the leaf shape, with the exception of decreased petiole length. The *tangled1* mutation disrupts the

spatial control of cytokinesis in the developing maize leaf (Smith et al., 1996; Cleary and Smith, 1998), and further work is needed to identify if the *lps1* mutation also affects the spatial control of cytokinesis.

One factor that was not measured in this study, but should be measured in subsequent research, is if there is a change in the amount and distribution of light penetrating into the soybean canopies of the *Lps1* and *lps1* NILs. Most light interception occurs at top and peripheral sections of the soybean canopy (Hatfield and Carlson 1978; Sakamoto and Shaw, 1967a and 1967b; Shaw and Weber, 1967). Sakamoto and Shaw (1967b) studied the distribution of light in the soybean canopy and determined that increases in yield could be achieved by breeding for lines that allowed deeper penetration of light into the canopy. This suggested method to improve soybean yield was verified by Kokubun (1988) who mechanically modified the canopy architecture to increase light penetration into the lower canopy and also verified by Johnston et al. (1969) who added supplemental light to the lower canopy which resulted in a 17% yield increase. In maize, Pendleton et al. (1968) demonstrated that increases in leaf erectness by mechanical and genetic means increased grain yield by 14% and 41%, respectively. Long et al. (2006) found that for canopies with an LAI of less than 2, horizontal leaves allow the greatest interception of daily incident solar radiation and canopies with greater LAI benefit from more vertically oriented canopies. Similarly, for canopies with greater LAI, Loomois et al. (1967) identified the ideal canopy as composed of horizontal leaves at the canopy bottom and increasingly more vertical leaf angles towards the canopy top. In soybean, the leaf angle can be affected by the length of the petiole. Hicks et al. (1969) observed that there was an increase in the petiole's angle of attachment to the main stem as the petiole length and total weight of the leaflets also

increased. This increase in the petiole angle of attachment was partially attributed to a decrease in light penetrating to the same level of the soybean canopy (Hicks et al., 1969) and suggests that *lps1*'s shorter petiole length could improve light infiltration into the canopy by decreasing the petiole's angle of attachment to the stem (Figure 29). Canopy light infiltration measurements need to be conducted to evaluate this hypothesis.

Seed weight was statistically decreased in the *lps1* NILs, which is important to note as seed weight is a component of grain yield calculations. However, if a fixed amount of energy is allocated to seeds, decreasing the seed size increases the number of seeds that can be produced and *vice versa*, so a decrease in seed size would not necessarily decrease yield as long as the number of seed is increased. In a study observing the changes that have occurred in soybean breeding across 58 years of cultivar releases, Morrison et al. (2000) found that cultivar seed weight is not correlated with cultivar year of release indicated that breeders are not specifically selecting for increased seed weight in order to increased grain yield. In fact, one study found that higher seed number is correlated with increased soybean grain yield (Rotundo et al., 2012).

One factor that was not measured in the *lps1* evaluations was if the decreased petiole length provides the plant a significant decrease in its carbon budget costs. The mass of petioles is considered small relative to the mass of the rest of the plant and is often not include in harvest index measurements due to this reason. However, the plant does expend resources to build and maintain the petiole, and thus it would make sense that a decrease in petiole length could decrease the plant's building and maintenance costs. Thus, it may be valuable to evaluate if the mutant's decreased petiole length results in measurable decreases in carbon budget costs, and if so, where are the resource savings

being alternatively allocated?

Preliminary yield trial results indicated that the *lps1* phenotype could potentially be valuable for improving soybean production through improved harvest index. Research by Morrison et al. (1999) found that there was a statistically significant linear increase of 0.47% in harvest index and an average yield increase of 0.5% per year with year of release in a survey of cultivars spanning 58 years of soybean improvement as those same cultivars displayed. The *lps1* mutant phenotype was found to significantly increase the relative mean harvest index, averaged across both locations, by 10.2% ($p = 0.0146$), while at the same time producing a relative mean increase in grain yield of 8.9%, although the grain yield increase was not found to be statistically significant ($p = 0.0696$). The St. Paul location showed higher *lps1* relative mean values for increased harvest index and grain yield as compared to the performance of the *lps1* NILs at the Rosemount location. The difference of performance may be due to management differences as the St. Paul location had better weed control (data not shown), and the St. Paul plots were not affected by pocket gophers as occurred at the Rosemount location. While preliminary data suggests that the *lps1* phenotype is not detrimental to yield, additional yield trials are necessary to validate if the *lps1* phenotype can be utilized for improving soybean yield through improved harvest index.

The short petiole phenotype of the *lps1* mutant produces a more narrow plant profile, and as a result, *lps1* lines need to be planted on narrow rows in order to be able to close the crop canopy. Currently, many farmers grow soybeans on 76cm wide rows as this allows the farmer to plant both corn and soybeans using the same planter. Thus, if

subsequent tests identify that the *lps1* mutant is beneficial to soybean production, farmers wishing to plant *lps1* lines will need to utilize narrower row spacings for their soybean production. The change in row spacings would require farmers to purchase a new planter and altered agronomic techniques. However, previous studies have demonstrated that planting soybeans on row spacings less than 76cm does provide multiple benefits. Narrower rows results in a more equidistant plant spacing which can lead to greater interception of the light and increase canopy leaf area development in the early season (Bullock et al., 1998; Weber et al., 1966; Yelverton and Coble, 1990), reduce weed growth and weed resurgence (Yelverton and Coble, 1990), and result in efficient use of water, light, and mineral nutrients during the season (Burnside and Colville, 1963; Andrade et al., 2002). All of these aforementioned benefits help to increase the yield potential of soybeans planted on narrow rows compared to wider row spacings (Ablett et al., 1991; Andrade et al., 2002; De Bruin and Pedersen, 2008; Bullock et al., 1998; Burnside and Colville, 1964; Cooper, 1977; Cooper, 1981; Cooper and Jeffers, 1984; Costa et al., 1980; Devlin et al., 1995; Elmore, 1998; Lueschen et al., 1992; Oplinger and Philbrook, 1992; Wax et al., 1977; Weber et al., 1966; Yelverton and Coble, 1991). Therefore, if future yield trials indicate that the *lps1* trait is beneficial for yield, farmers that plant *lps1* soybean lines in narrow row spacings could see increased yields from both the *lps1* trait and from narrow row plantings as compared to planting wild-type soybeans on wide row spacings.

CHAPTER 5

Future Directions

This dissertation has demonstrated that new genomic technologies have allowed for the improvement of biparental mapping, through increased marker resolution; however, it may be possible to employ improved population designs and experimental designs to more fully leverage the new capacity of these new genomic technologies.

Currently, chemical or radiation induced mutagenesis often involves making mutations in only one genetic background. Next, for a forward genetics approach, the population is phenotypically screened for one or more traits of interest. After a mutant of interest is found, the researcher outcrosses this single mutant to a wild-type line to generate the segregating population. This single outcross mutant population is then phenotyped, and the polymorphisms created by outcrossing are used for mapping. A significant amount of resources might be employed to genotype and advance a large number of individuals in this population. However, it may be possible to change the design of the experiment to increase the efficiency by halving the number of lines developed per trait mapped without reducing the number of recombinants. For example, if two chemical or irradiation populations were developed using two genetically distinct genetic backgrounds, perhaps each half the size of a single population, then a mutant identified in one population could be crossed to a mutant identified in the second population in order to develop a single mapping population capable of mapping both traits. This scenario would require that both traits of interest would not interfere with each other and could be mapped simultaneously. This approach would likely be better

employed for chemical rather than irradiation induced mutants as irradiation induced genome rearrangements in one mutant could interfere with the ability to map the trait from other mutant. While the combined mapping (Co-Map) approach would add additional complexity, this mapping method has the benefit of increased efficiency in terms of lines developed per trait mapped.

A variant of the aforementioned mapping of multiple traits could be done using a mutant in which multiple mutant traits have been stacked. For instance, certain individual lines in the Soybean Isoline Collection contain several mutant traits (Bernard et al., 1991). By out crossing just one of these individuals containing multiple mutant alleles to a wild-type line contrasting for these traits, multiple mutant traits could be mapped in the same population.

There are other potential improvements that could be made on the genotyping side. An additional method to consider for mapping traits is to use low depth whole genome resequencing on each individual of a mapping population. If each individual's sequence is fitted with unique library barcodes and then sequenced, individual recombination points could be tracked in each line. The mapping population could be synthetically bulked again and again for as many traits as are phenotyped. As each individual's recombination points can be determined, a narrow mapping interval and indeed the causative polymorphism can be determined. However, this method is currently more expensive and has little benefit over simply resequencing the crossed parents and conducting high density genotyping on the progeny. Thus it is unlikely that such a method will be used on bi-parental mapping populations until resequencing costs are

considerably reduced.

While, the use of whole genome resequencing of individual bi-parental mapping population lines is not currently cost effective, such a strategy would be very useful for certain next-generation population designs. The Multi-parent Advanced Generation Inter-Cross (MAGIC) next generation population design generates increased mapping resolution and improved statistical power for determining trait effects (Darvasi and Soller, 1995; Mott et al., 2000; Rockman and Kruglyak, 2008; Macdonald and Long, 2007). This population design calls for the intermating of multiple founding parents followed by several rounds of intermating the progeny of the initial crosses, and thus the design captures an allelic series across multiple genetic backgrounds. Fully realizing the benefits of increased mapping resolution and statistical power is predicated on obtaining sufficient marker density to identify all of the recombination events and segregating haplotypes. Whole genome resequencing of each population individual would permit the identification and tracking of all haplotypes with a high level of confidence to increase the accuracy of trait mapping and estimates of quantitative trait loci effects.

The preceding sections have discussed how current technologies have allowed researchers to consider and conduct new experimental designs. As new technologies arise, it will be important for researchers to wisely decide how and when to use these new technologies. In addition, scientists need to be inventing new experimental and population designs in order to fully employ the capabilities provided by new technologies.

ILLUSTRATIONS

FIGURES

Chapter 1 Figures

(No figures)

Chapter 2 Figures

Figure 1. Phenotypic evaluation of chlorophyll deficiency in the MinnGold mutant. (A) Visual comparison of the MinnGold mutant (left) versus the wild-type cultivar Williams 82 (right). (B) Total leaf tissue chlorophyll levels in the MinnGold, MinnGold non-transgenic segregants, Transgenic MinnGold T₁, and Williams 82.

A



B

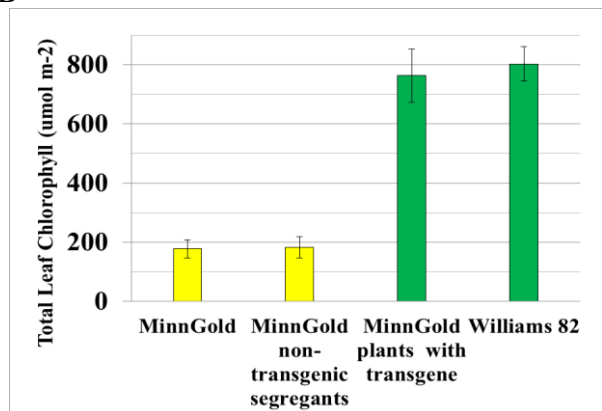
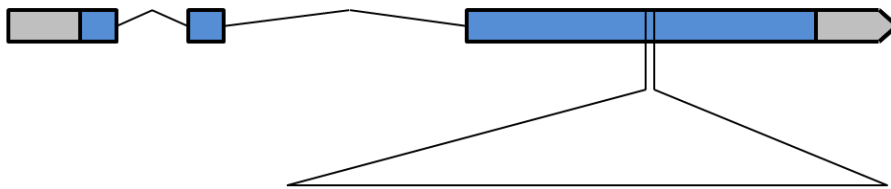


Figure 3. The *y11-2* and *y11* mutations in the candidate gene Glyma13g30560 appear to be novel *de novo* mutations.

(A) Gene diagram for the candidate gene Glyma13g30560: the 3 exons are indicated in blue, 5' UTR and 3'UTR are in grey. (B) Sequence results of MinnGold (*y11-2*) and T219 (*y11*) compared to a panel of diverse soybean lines. Of the 29 diverse soybean lines sequenced (15 are shown), only *y11-2* was found to have Adenine rather than Guanine at position 33146523 and only *y11* was found to have the Guanine rather than the Adenine at position 33146529. The *y11-2* SNP causes a nonsynonymous change from Arginine to Glutamine, and the *y11* SNP causes a nonsynonymous change from Glutamine to Arginine.

A Glyma13g30560



B

	Gm13:33,146,509..33146523..33146529..33,146,538	
Consensus	AGAAGGGGAGCTCC G GCCGC A GCTGCTAGA	
MinnGold (<i>y11-2</i>) A A	
T219 (<i>y11</i>) G G	
Wm82 G A	
Archer G A	
Minsoy G A	
Noirl G A	
Forrest G A	
Hendricks G A	
Missoy G A	
Parker G A	
Bert G A	
Anderson G A	
Dunfield G A	
Lincoln G A	
Odgen G A	
Perry G A	
Richland G A	
	Arginine CGG	CAG Glutamine
	Glutamine CAG	CGG Arginine

Figure 4. A Glyma13g30560 wild-type transgene complements the chlorophyll deficiency phenotype.

The top panel shows the phenotype of 17 T₁ plants segregating for the presence of the Glyma13g30560 wild-type transgene. NT indicates the no template negative control. MG indicates untransformed MinnGold. The perfect correlation between the presence of the transgene and wild-type phenotype in the segregating T₁ progeny indicates that the transgene is restoring wild-type function.

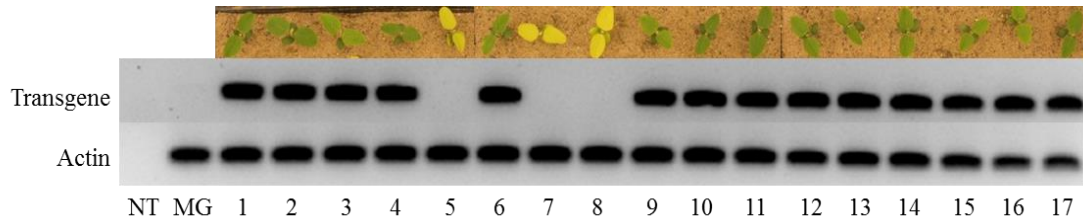


Figure 5. Phenotypic classes for chlorophyll deficiency mutants.

(A) Leaves of *Y11/Y11*, *Y11/y11*, and *y11/y11* leaves showing the distinctive phenotypic classes. (B) Leaves of WT/WT, WT/CD-5, and CD-5/CD-5 leaves showing the distinctive phenotypic classes.

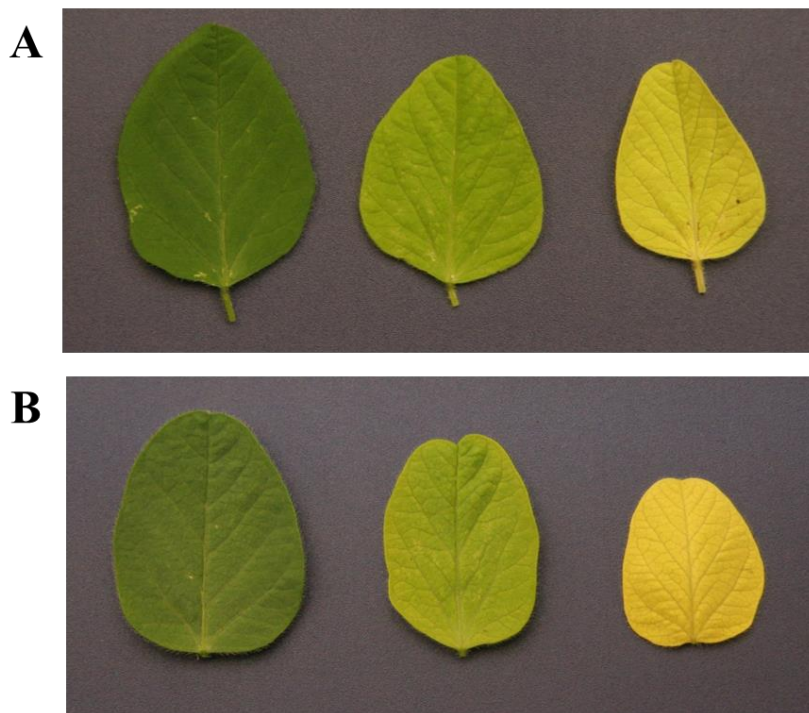


Figure 6. *y11* CAPS assay.

A Cleaved Amplified Polymorphic Sequences (CAPS) assay of nineteen individuals segregating for the presence of the candidate *y11* SNP. In the rightmost lanes, D indicates a *y11/y11* digested sample and a ND indicates a *y11/y11* sample that was not digested. The perfect cosegregation of the candidate *y11* SNP with the foliage phenotype provides additional information to suggest that the candidate SNP is responsible for the chlorophyll deficient phenotype.

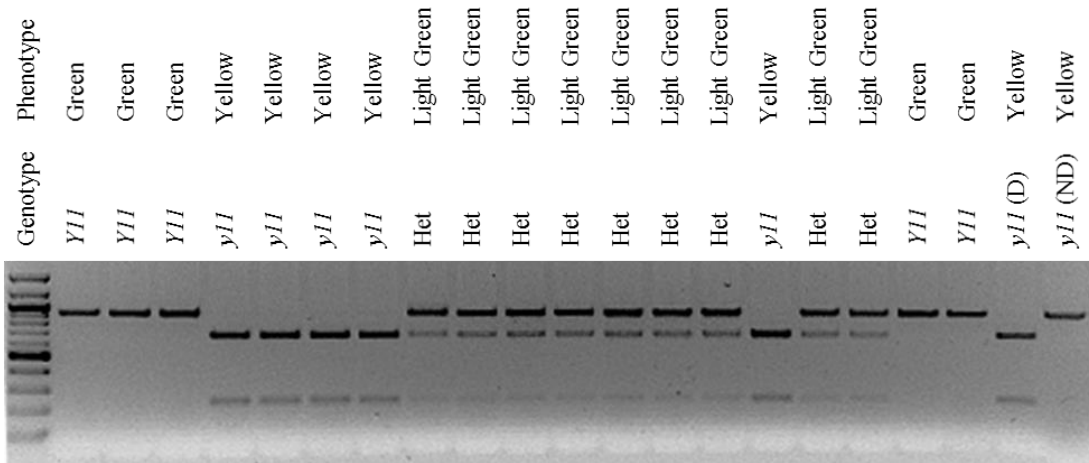


Figure 7. A Cleaved Amplified Polymorphic Sequences (CAPS) assay of seventeen individuals segregating for the presence of the candidate CD-5 SNP.

In the rightmost lanes, D indicates a WT/WT digested sample and a ND indicates a WT/WT sample that was not digested. The perfect cosegregation of the candidate CD-5 SNP with the foliage phenotype provides additional information to suggest that the candidate SNP is responsible for the chlorophyll deficient phenotype.

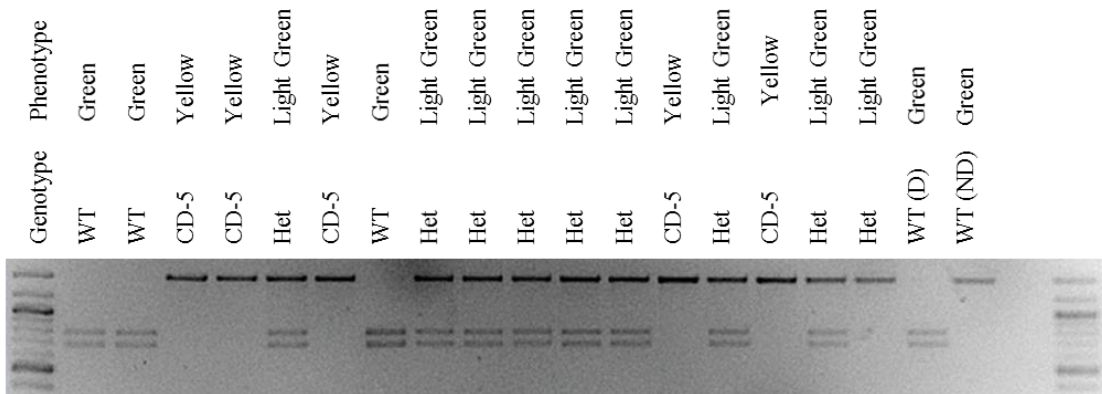


Figure 8. The CD-5 mutation in the candidate gene Glyma15g08680 appears to be a novel *de novo* mutation.

(A) Sequence results of CD-5 compared to a panel of diverse soybean lines. Of the 29 diverse soybean lines sequenced (15 are shown), only CD-5 was found to have Guanine rather than Adenine at position 6137419. The CD-5 SNP causes a nonsynonymous change from Glutamine to Arginine. (B) Sequences from Glyma13g30560 and Glyma15g08680 showing that the *y11* and CD-5 SNPs in the separate genes occur at the same relative SNP position, but that sequences from the two genes can be differentiated by nearby SNPs. Sequences for Glyma13g30560 and Glyma15g08680 are from positions Gm13:33,146,509..33,146,538 and Gm15:6,137,439..6,137,410, respectively.

A	Gm15:6,137,439... 6137419 ...6,137,410
Consensus	AGAAGGTGAGCTCCGGCCAC A GCTGCTGGA
CD-5 G
Wm82 A
Archer A
Minsoy A
Noirl A
Forrest A
Hendricks A
Missoy A
Parker A
Bert A
Anderson A
Dunfield A
Lincoln A
Odgen A
Perry A
Richland A
	CAG Glutamine
	CG G Arginine
B	
Glyma13g30560	AGAAGGGGAGCTCCGGCC G CAGCTGCT A GA
Glyma15g08680	AGAAGG T GAGCTCCGGCC A CAGCTGCT G GA
SNP differences*.....*.....*.....
Glyma13g30560 <i>y11</i>G.....G GA..
Glyma13g30560 <i>y11-2</i>G..... A ...G.....A..
Glyma15g08680 CD-5T.....A GG..

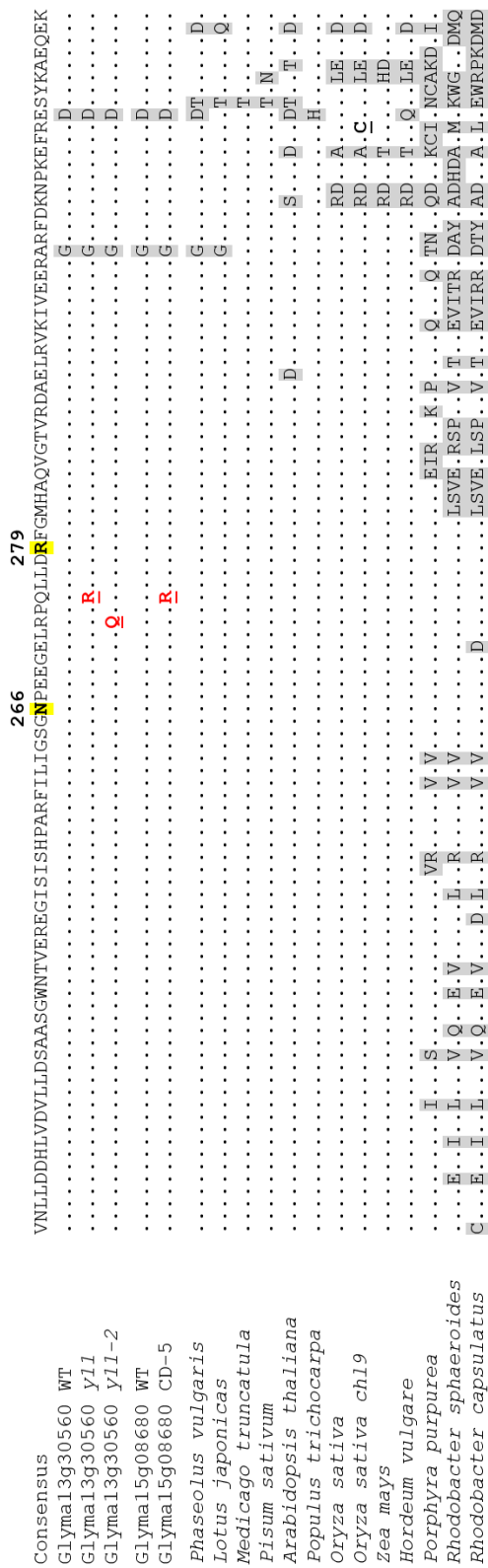


Figure 9. Polypeptide sequence alignment of the Mg-chelatase ChII subunit for the interval surrounding the *y11*, *y11-2*, and CD-5 mutations. The shaded 266N and 279R nucleotides indicate the amino acids predicted to be involved in an ATP binding domain and an Arginine binding domain, respectively (Marchler-Bauer, 2013). The bolded 273Q and 275R denote the *y11-2* and *y11* non-synonymous amino acid changes in Glyma13g30560, respectively. The bolded 275R denotes the CD-5 non-synonymous amino acid change in Glyma15g08680. The shaded residues indicate residues that are not completely conserved across species. The wild-type soybean sequence and the sequences of the other species, listed in order, are available in the NCBI database under accession numbers: XP_003543008, XP_003546019, XP_007148073, AFK38677, XP_003593716, AET86637, NP_193583, XP_002316838, NP_001050493, ACN32024, AAA99720, U38804, AF017642, and Z111165.

Figure 10. Amino acid sequence comparison of ChIIa (Glyma13g30560) to ChIIb (Glyma15g08680)

The amino acid sequence comparison of Glyma13g30560 and Glyma15g08680 shows the high degree of similarity between the two Mg-chelatase subunits. The two boxed residues indicate the positions of the *y11-2* (R273Q) mutation and *y11* and CD-5 (Q275R) mutations.

Consensus	MAS.LGTSSI	AVLPSR..SS	.SSKPSIHTL	SLTSGQ.YGR	KFYGGIGIHG	50
Gma ChIIa	...A.....YF..	S.....N...	
Gma ChIIb	...T.....CI..	F.....S...	
Consensus	IKGR.QLSV.	NVATEVNSVE	QAQSIASKES	QRPVYPFSAI	VGQDEMKLCL	100
Gma ChIIaA....T	
Gma ChIIbS....A	
Consensus	LLNVIDPKIG	GVMIMGDRGT	GKSTTVRSLV	DLLPEIKVVA	GDPYNSDPQD	150
Gma ChIIa	
Gma ChIIb	
Consensus	PEFMGVEVRE	RVLQGEELSV	VLTKINMVDL	PLGATEDRVC	GTIDIEKALT	200
Gma ChIIa	
Gma ChIIb	
Consensus	EGVKAFEPGL	LAKANRGILY	VDEVNLLDDH	LVDVLLDSAA	SGWNTVEREG	250
Gma ChIIa	
Gma ChIIb	
Consensus	ISISHPARFI	LIGSGNPEEG	ELR EQ LLDRF	GMHAQVGTVR	DAELRVKIVE	300
Gma ChIIa	
Gma ChIIb	
Consensus	ERGRFDKNPK	EFRDSYKAEQ	EKLQQQITSA	RSVLSSVQID	QDLKVKISKV	350
Gma ChIIa	
Gma ChIIb	
Consensus	CAELNVDGLR	GDIVTNRAAK	ALAALK.RD.	VSAEDIATVI	PNCLRHLRKL	400
Gma ChIIaG..N	
Gma ChIIbE..K	
Consensus	DPLESIDSGL	LVTEKFYEVF	S	421		
Gma ChIIa			
Gma ChIIb			

Figure 11. Speculated combinations of *y11-2* (A), *y11* (B), and CD-5 (C) mutant and wild-type ChIIa and ChIIb subunits arranged into hexameric rings. The resulting chlorophyll phenotypes from the various combinations of wild-type and mutant proteins are indicated by the dark (dark green), lighter (lighter green), and lightest (yellow) background colors. ChIIa and ChIIb subunits are depicted in white and grey colors, respectively. Solid lines around the subunits indicate wild-type proteins while dashed lines around the subunits indicate mutant proteins. A solid (blue) circle in the center of the hexamer indicates the hexamer is capable of wild-type activity.

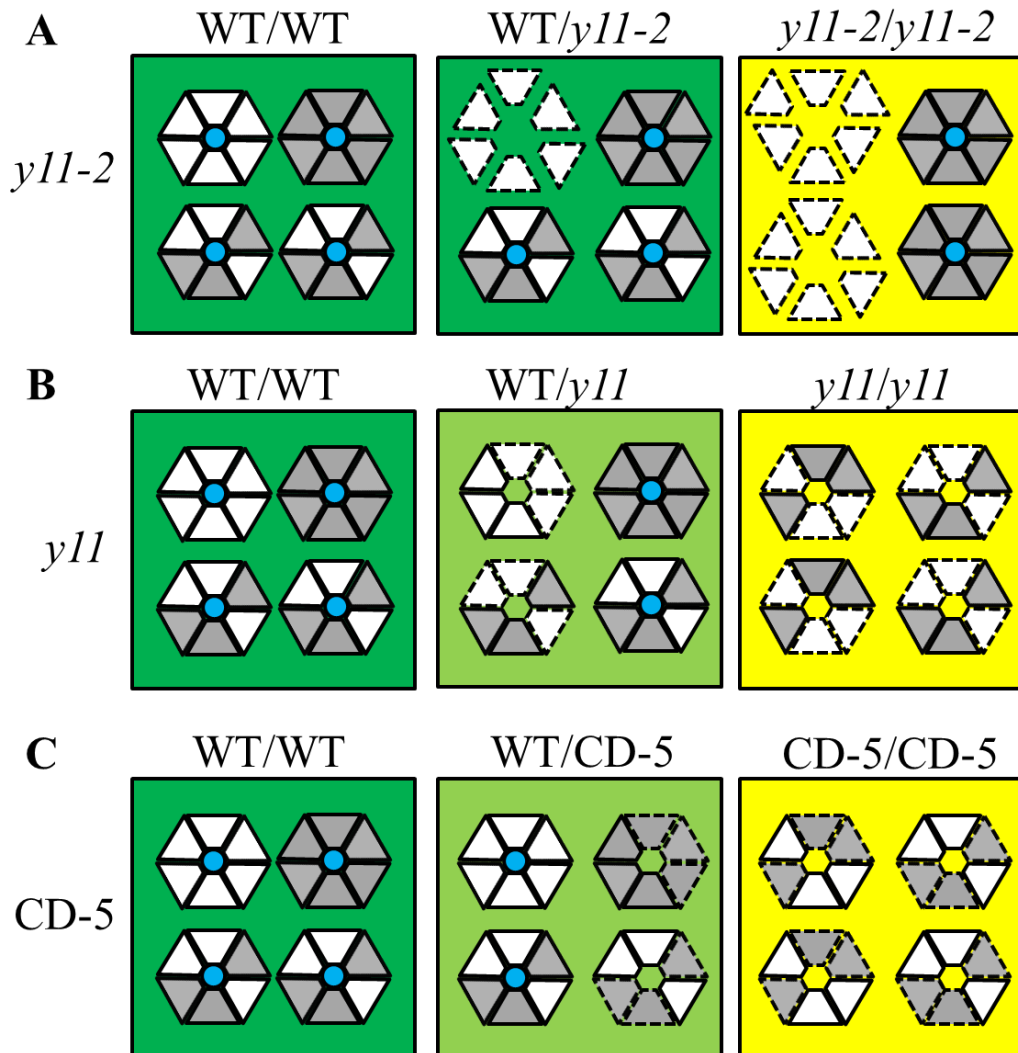
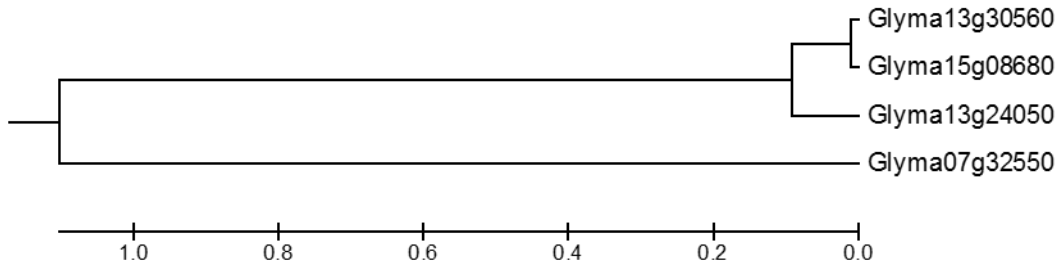


Figure 12. The inferred evolutionary history for the four Glycine Max ChII subunits calculated using the UPGMA method.

The UPGMA neighbor joining tree shows the close sequence similarity of Glyma13g30560 with Glyma15g08680. The optimal tree with the sum of branch length = 2.30757376 is shown.

UPGMA neighbor joining tree for the four *Glycine Max* ChII Subunits.



Chapter 3 Figures

Figure 13. The mutated Glyma.20G019300 allele co-segregates with the *gnarled* phenotype.

(A) Three primers were used to generate a co-dominant marker that differentially amplifies wild-type and mutant alleles. The arrows indicate both the position and the direction of the primers B121R, B124R, and B124F. The B124F and B124R primers amplify a 708 bp fragment from the wild-type allele, and the B121R and B124R primers amplify a 188bp fragment from the mutant allele. The combination of the inversion and deletion in the mutant allele orients the B124R primer such that it can amplify a fragment when paired with the B121R primer. (B) Perfect co-segregation was observed between the phenotypic classes and the expected genotypic classes among a population of 50 F₃ individuals. The parent lines (R55C01 and 'Noir 1') are shown. Mutant (M) individuals exhibited only the 188 bp fragment, and wild-type (Wt) individuals exhibited either both fragments (heterozygous (Het)) or only the 708 bp fragment

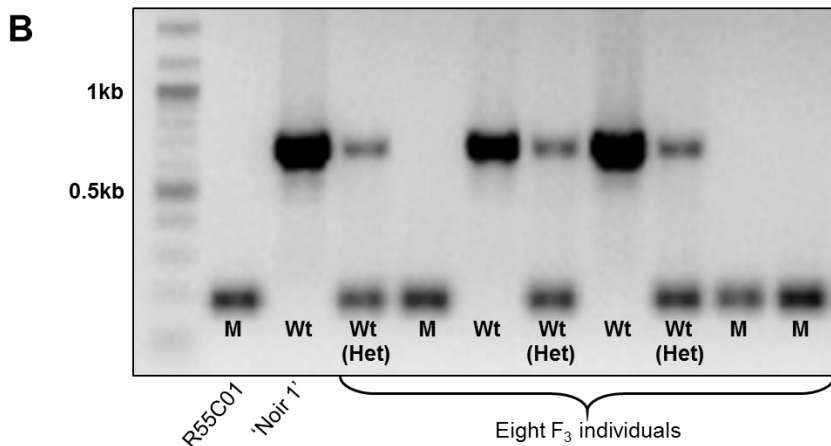
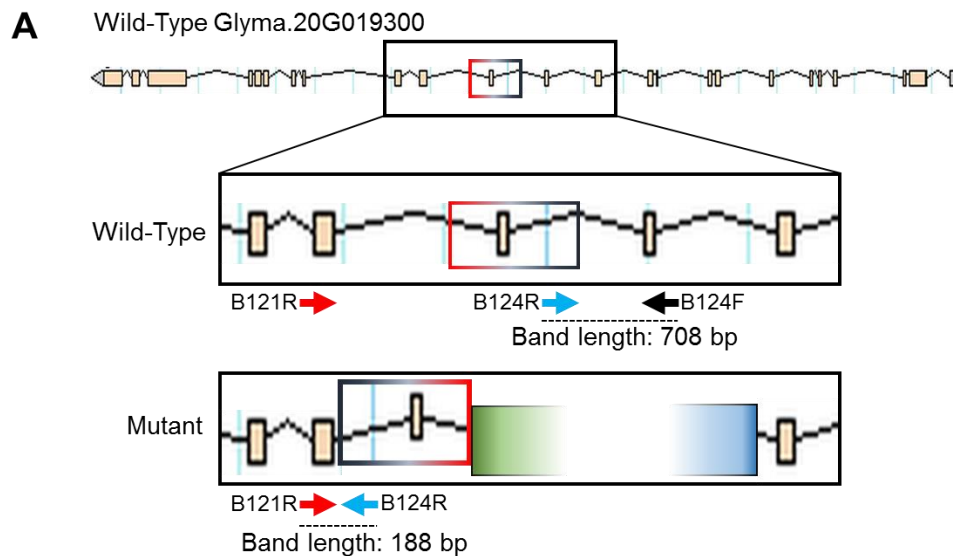


Figure 14. Phenotype and course genetic mapping of the *gnarled* trichome mutant. (A) Phenotypes of mutant (left) and wild-type (right) shoots. SEM leaf trichome images of the *gnarled* mutant R55C01 (B) and the wild-type line ‘M92-220’ (C). The mutant trichomes are wide, short, flaccid, and lay on the surface of the leaf or stem, as compared to the wild-type trichomes (narrow, long, straight, and project outward from the leaf or stem). Scale bars in (B) and (C) are 1 mm. (D) BSA-WGS allele frequencies were calculated for F₂ bulked samples that consisted of 50 mutant (dark color line / red line) and 50 wild-type (light color line / blue line) individuals. The allele frequencies were calculated as the proportion of reads containing the wild-type parental SNP (i.e. SNPs that match the wild-type parent ‘Noir 1’) at each position for over 16,000 polymorphic SoySNP50K positions. The allelic frequencies are shown as the average value across a 21 SNP sliding window. The obvious spread in allele frequencies indicates that the causative locus is located on chromosome 20.

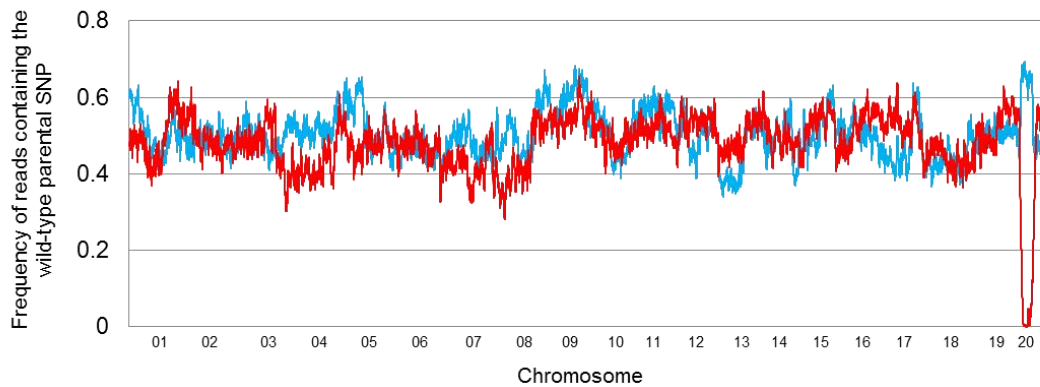
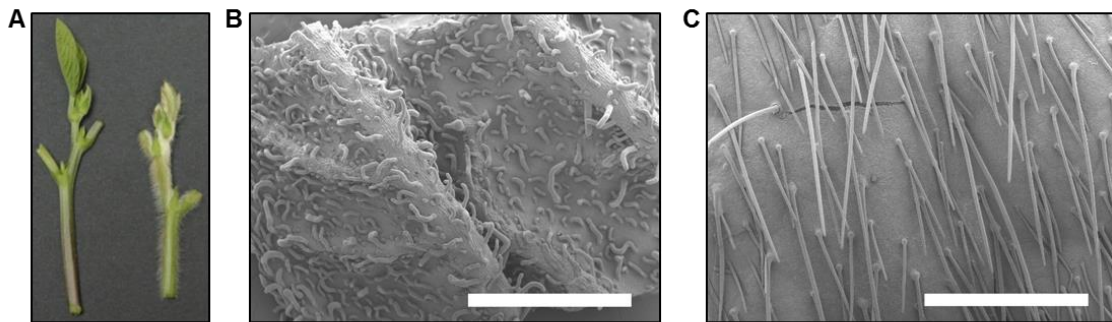
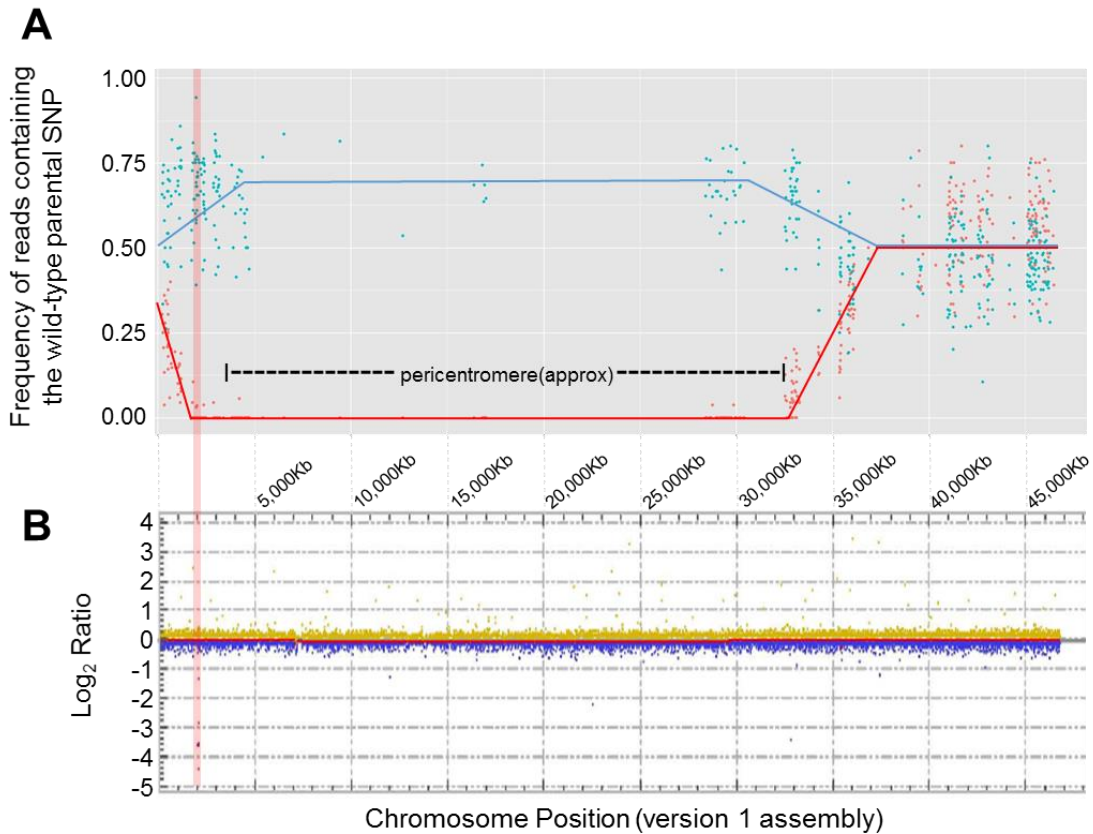


Figure 15. Genetic mapping of the *gnarled* mutant and physical mapping of the deletion on chromosome 20.

Coincidental mapping of the (A) WGS-BSA mapping interval with (B) a deletion detected by array Comparative Genomic Hybridization (aCGH). In (A), light color (blue) data points indicate the ‘Noir 1’ SNP frequency at each marker position in the wild-type bulk; dark color (red) data points indicate the ‘Noir 1’ SNP frequency at each marker position in the mutant bulk. Light (blue) color and dark (red) color lines, respectively, are drawn in (A) to assist in visualizing the separation in the bulk allele frequencies. A lightly shaded vertical box shared between (A) and (B) identifies the chromosome region containing the *GmNAP1* gene. In (b), probes below the 0.0 \log_2 value indicates the absence of mutant DNA (i.e. a putative deleted segment)



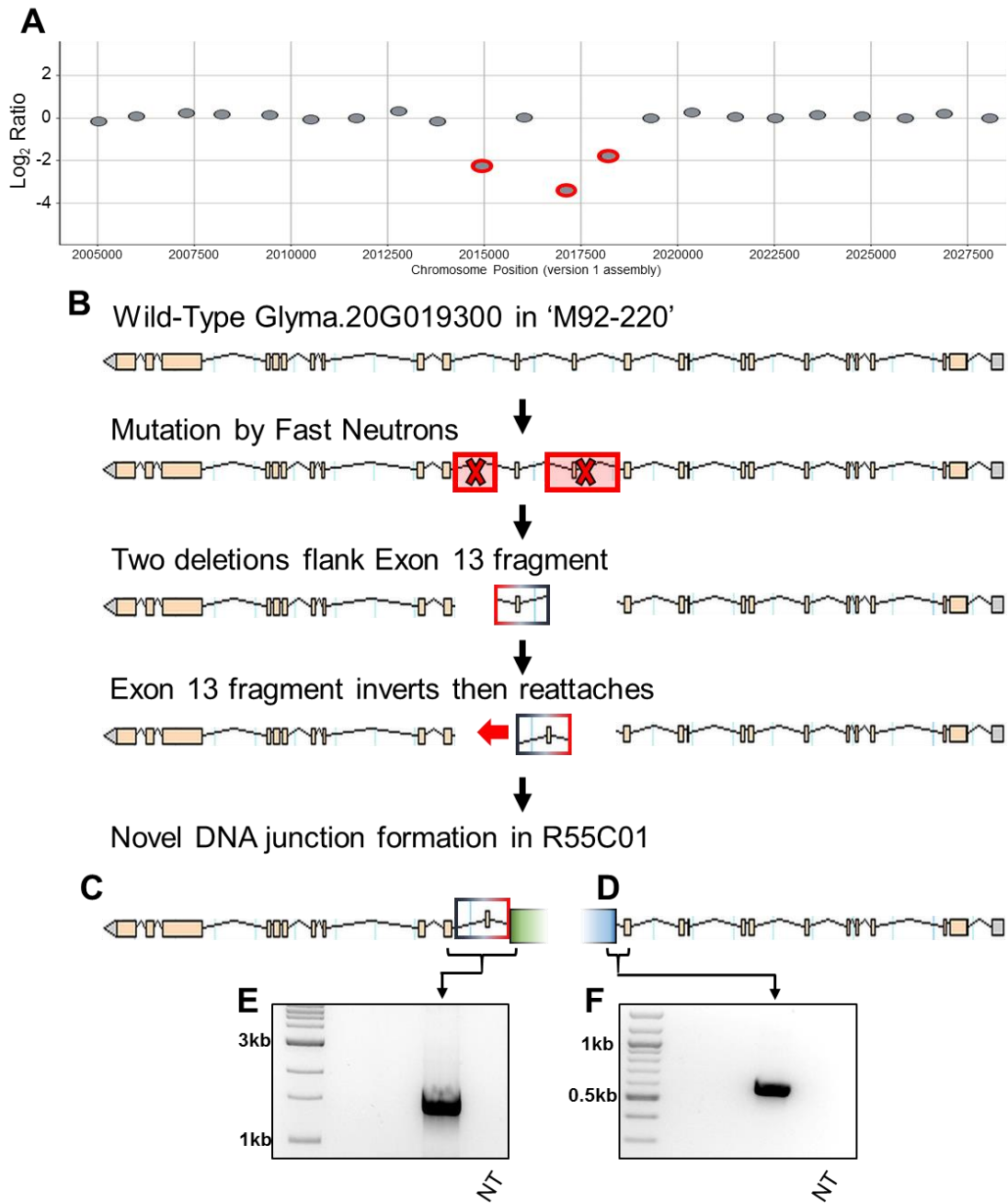


Figure 16. Mutations in the candidate gene demonstrate the complexity of mutations that can occur by fast neutron mutagenesis.

(A) aCGH report depicting two deletions in *GmNAP1* indicated by probes with corrected \log_2 ratios of less than -2. aCGH array was designed using version 1 assembly (Glyma.Wm82.a1.v1.1), thus v1 positions are listed. (B) Wild-type Glyma.20G019300

and fast neutron mutations to the gene. (C) The inverted 13th exon connected to the second half of the gene forms a novel junction between Gm20:2,010,290 to Gm20:2,007,928 (positions are according to the version 2 genome assembly, Glyma.Wm82.a2.v1). The Gm20:2,009,152 side of the inverted fragment is connected to sequence found at Gm20:16,920,485. (D) The first half of the gene is interrupted at Gm20:2,012,311 and is connected to sequence found at Gm20:16,939,673. A novel 22bp sequence was found in the junction. PCR amplification was used to confirm the novel DNA junctions created by fast neutron mutagenesis. (E) A 1.4 kb fragment spanning across two novel junctions created in the second half of the gene. (F) A 605 bp fragment spanning the novel junction created in the first half of the gene. The samples tested were (left to right): ‘Williams 82’, ‘M92-220’, R55C01, and a no template control. The orientations of the sequences at the junctions do not suggest that a single contiguous piece was inserted into Glyma.20G019300, and the extent of chromosomal rearrangements that occurred on chromosome 20 is unclear at this time

Figure 17. RNA-seq read alignment density for each exon of Glyma.20G019300 in wild-type and *gnarled* mutant plants.

RNA-seq reads mapping to Glyma.20G019300 clearly illustrate the lack of transcription from exons 12-14 in *NAP1* mutant plants. The height of the histogram indicates read depth along the length of the entire gene, with transcription peaks corresponding to exon sequences. Colored bars indicate SNPs relative to the ‘Williams 82’ reference genome sequence. Transcription of exon 11 appears to be up-regulated in mutant tissues compared to the wild-type. Exons 12, 13, and 14 are transcribed in all tissues of the wild-type plant but are not transcribed in the mutant plant, corresponding to the fast neutron induced deletions and structural rearrangements. Transcription of exons 15-23 is generally lower in tissues from the mutant plant compared to the wild-type plant

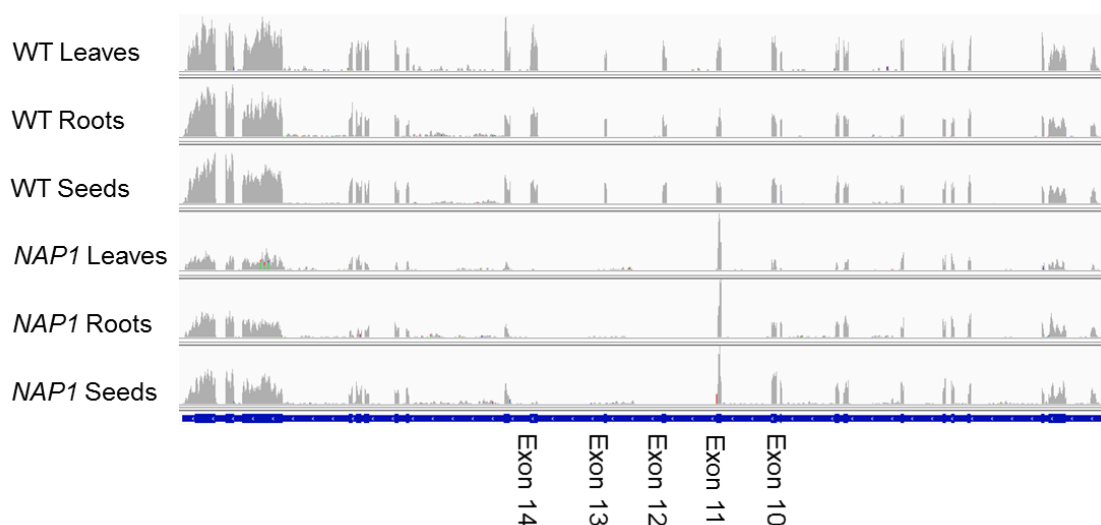


Figure 18. Soybean *GmNAP1* functionally complements *Arabidopsis nap1* mutant (*grl-4*).

(A) SEM image of wild-type trichomes on a Col-0 leaf. (B) SEM image of *gnarled* trichomes on a *nap1* mutant (*grl-4*). (C) SEM image of wild-type trichomes on a T₂ *grl-4* plant complimented with the soybean *GmNAP1* transgene. Scale bars in (A-C) are each 200 μ m. (D) Amplification of *GmNAP1* transgene in 20 T₁ *Arabidopsis grl-4* individuals with wild-type trichomes confirms that the *GmNAP1* is able to functionally compliment the *Arabidopsis nap1* mutant. From the left: soybean cv. 'Williams 82', *Arabidopsis nap1* mutant (*grl-4*), 20 *Arabidopsis grl-4* mutants transformed with the *GmNAP1* transgene and displaying a wild-type trichome phenotype. The fragment amplified spans from the promoter region into the first exon. The band size of 'Williams 82' is 548 bp, and the band size of the 20 *Arabidopsis* individuals is 556 bp. The difference of 8 bp is due to the insertion of an *AscI* restriction site in the *GmNAP1* transgene construct, just upstream of the ATG start site, which was added during construct assembly

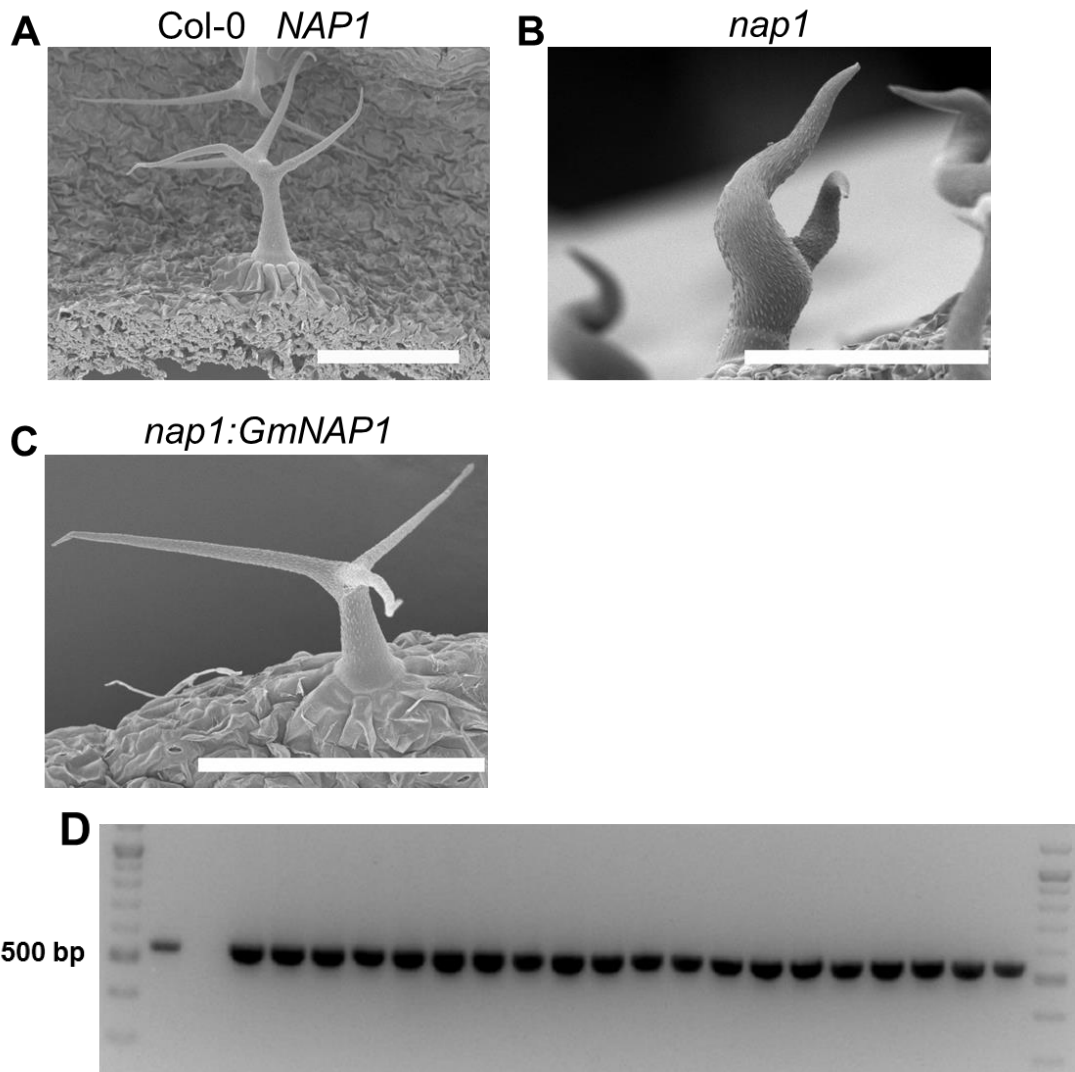


Figure 19. *p2* introgression interval identified on chromosome 20.

Positions are given for polymorphic SoySNP50K markers in the genomic region. The *p2* allele, from the donor line T31, was previously backcrossed into the recurrent parents ‘Clark’ and ‘Harosoy’ to generate five Near Isogenic Lines (NILs) as part of the Soybean Isoline Collection. The donor parent genotypes are colored in the dark color (red), recurrent parent genotypes are slightly less dark color (blue), heterozyous (Het.) genotypes are in a mix of the donor and recurrent parent colors (red and blue), and missing genotypes are color coded in light grey. Heterozyous scores are likely due to heterogeneity in the NIL. Examination of the five *p2* NILs’ genotypes identified a single introgression interval (566 kb) in the genome in which all five lines shared the donor parent genotype (Gm20:1,582,950-2,148,735). This interval contains the *GmNAP1* gene (Gm20:1,999,216-2,021,765).

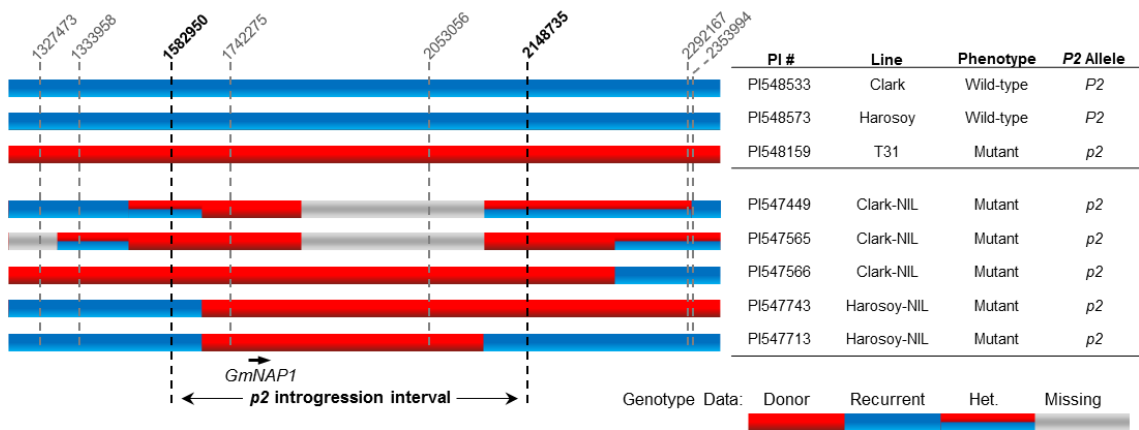


Figure 20. Sequence comparison of T31, ‘Williams 82’ (Wm82), and 25 diverse wild-type lines for the interval flanking the frame shift deletion found in T31. Sequencing of Glyma.20g019300 (*NAPI*) exon 22 in 25 diverse wild-type lines identified that the 1 bp deletion is unique to T31. Sequence differences between the two soybean *NAPI* paralogs (highlighted in grey) make it possible to differentiate the chromosome 20 (Gm20) and chromosome 7 (Gm07) sequences. The given sequence positions are for genome assembly version 2 (Glyma.Wm82.a2.v1). T31’s 1 bp deletion is highlighted in gray. The resulting frame shift caused an early stop codon starting 64 bases downstream from the single base deletion. The early stop codon is shown in bold in the T31 downstream sequence

Wm82	Gm07	39610713	ACATGCTAGTGAC AG AATGGATCCT GA ATTATCAAGTC	39610676
Wm82	Gm20	2000396	ACATGCTAGTGAGGGGAATGGATCCCGAATTATCAAGTC	2000359
T31			ACATGCTAGTGAGGGGAATG-ATCCCGAATTATCAAGTC ... (45bp) ... TGA	
Archer		G.....	
Minsoy		G.....	
Noirl		G.....	
Missoy		G.....	
Parker		G.....	
PI468922		G.....	
Bert		G.....	
A.K. (Harrow)		G.....	
Captial		G.....	
CNS		G.....	
Dunfield		G.....	
Illini		G.....	
Jackson		G.....	
Lincoln		G.....	
Forrest		G.....	
Lee		G.....	
Odgen		G.....	
Perry		G.....	
Richland		G.....	
Roanoke		G.....	
S-100		G.....	
Haberlandt		G.....	
Liu yue bao		G.....	
Shirome		G.....	
Zontanorukon		G.....	

Chapter 4 Figures

Figure 21. Short petiole phenotype of *lps1*.

(A) The five uppermost consecutive petioles from single F_{5,6} NIL plants demonstrating that the *lps1* mutation shortens petiole length. The mutant *lps1* petioles are on the left, and wild-type *Lps1* petioles are on the right. The reference bar is 10cm. Cell profiling of the mutant (B) and wild-type (C) petioles identified that the mutant petioles exhibit a higher frequency of collapsed or undeveloped cells (see arrows) as compared to wild-type.

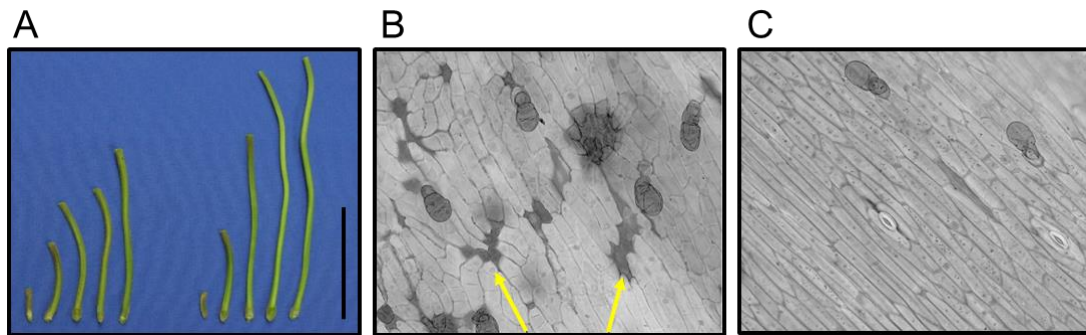


Figure 22. F_{2,3} *lps1* Bulked Segregant Analysis.

Bulked Segregant Analysis results indicated that the *lps1* mutation is located at the bottom end of chromosome 16.

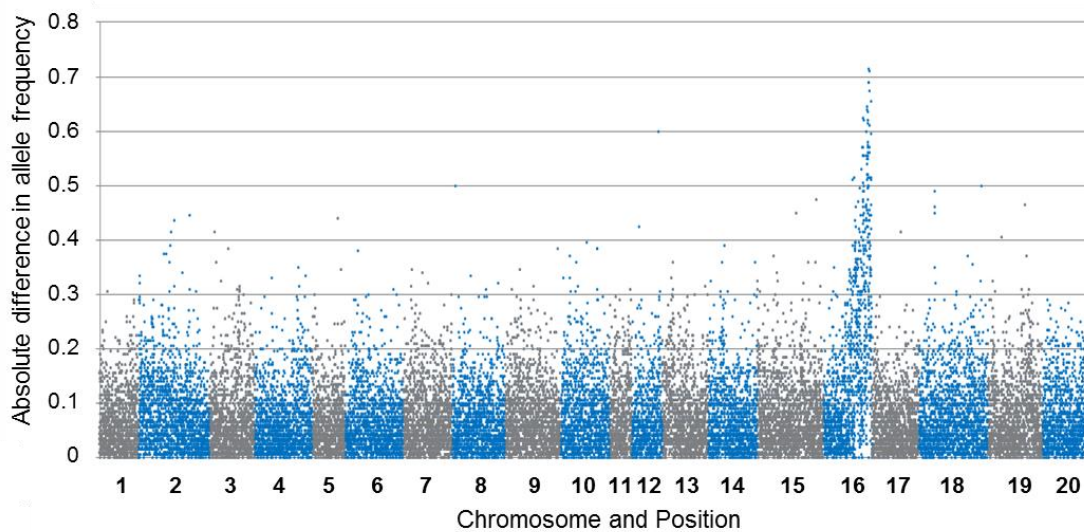


Figure 23. *lps1* Fine mapping interval and candidate gene.

(A) Fine mapped interval for *lps1*. An 84.9 kb interval (Gm16: 36831753-36916624) was found to co-segregate with the *lps1* phenotype. This interval contained 11 genes. The 84.9kb sequence in D76-1609 was compared against ‘Williams 82’ and seven wild-type lines to identify polymorphisms unique to D76-1609. This led to the identification of 8 polymorphisms in the interval. As D76-1609 occurred as a spontaneous mutation, the presence of these polymorphisms were tested on the parents of D76-1609 to see if any of the polymorphisms were unique to D76-1609. SNP changes with their polymorphic options, deletions are listed at “X” with a subscript depicting their allele type, and a small insertion is listed as (T/TAAG). Only polymorphism #2, an in-frame insertion of AAG in Glyma.16g209100 (filled in arrow) was found to be unique to D76-1609, as all of the other polymorphisms were found in one or more of the parents of D76-1609.

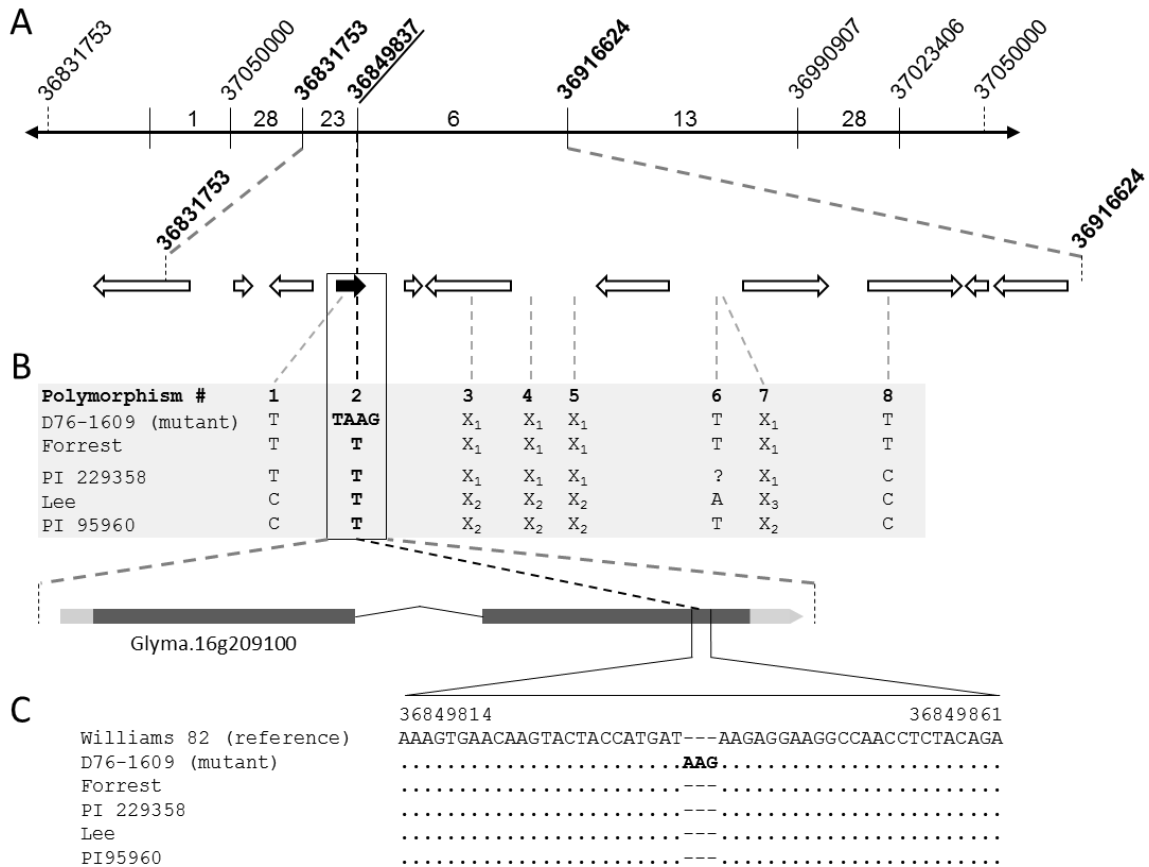


Figure 24. Uniqueness of the candidate *lps1* mutation in soybean germplasm. The four wild-type parents of D76-1609 and an additional 29 wild-type soybean lines (27 shown) were checked for the presence of the AAG insertion found in the mutant line. Results indicate that the polymorphism is unique to the mutant line D76-1609 and agrees with the spontaneous occurrence of the mutant phenotype. The AAG insertion results in the addition of an *Ear1* restriction site that when digested produces a mutant 383bp and 70bp bands versus the 453bp band for the wild-type lines as shown in the CAPS assay shown above. Lane 1: no template control; lane 2: uncut control; lane 3: mutant band at 383bp; lanes 4-7: the four parents of the mutant line: PI 229358, PI 95960, Forrest, and Lee with the 453bp wild-type band; and lanes 8-34: 27 wild-type lines also with a 453bp wild-type band.

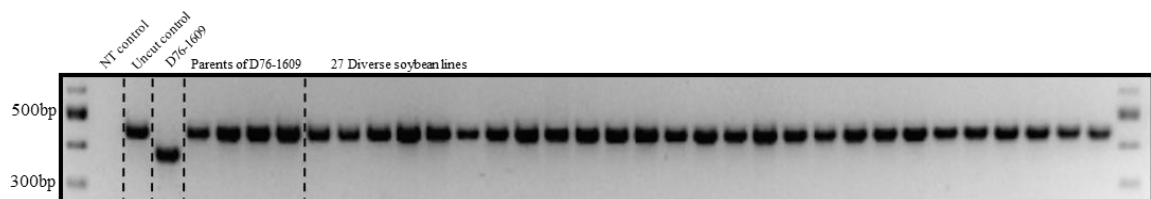


Figure 25. Petiole length is affected by environmental conditions. An individual with primarily long petioles and is homozygous for the *lps1* AAG insertion displayed the short petiole phenotype on only a single petiole on the plant (see arrow), suggesting that environmental conditions may affect the expressivity of the short petiole phenotype.



Figure 26. Expression differences between the *lps1* candidate gene (Glyma.16g209100) and its paralog (Glyma.09g159900). Expression difference of (A) Glyma.16g209100 and (B) Glyma.09g159900.

A



B



Figure 27. Preliminary yield trial results for short petiole versus long petiole Near Isogenic Lines.

Twelve NIL pairs were planted in 10-inch row yield plots at two locations in one year with some lines only having one replication per location. Average performance differences between the short and long petiole lines are given in percentages. For example, on average the short petiole lines had 8.9% more grain yield than long petiole lines. Error bars displayed are using standard error. These preliminary yield trial results indicate that the *lps1* short petiole trait is not detrimental to yield and may be beneficial through improving harvest index.

Preliminary Yield Trial for Short versus Long Petiole NIL Pairs
Average % advantage of short petiole lines over long petiole lines

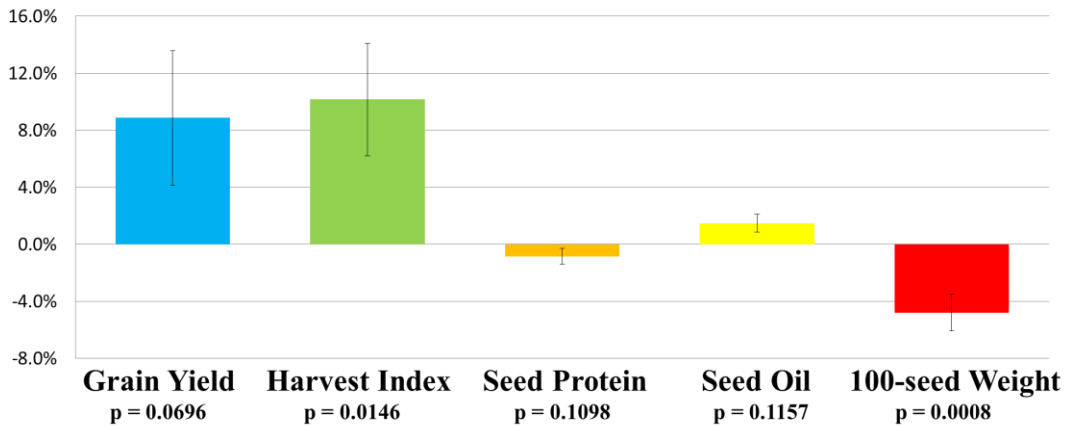


Figure 28. Petiole lengths and cell lengths from paired Near Isogenic Lines. (A) Average petiole lengths from five separate $F_{5:6}$ NIL pairs indicating that the *lps1* mutation reduces petiole length by about 40%. (B) Comparisons of the average petiole epidermal cell lengths from five $F_{5:6}$ NIL pairs indicate that the reduction in cell length is a major contributor to the observed differences in petiole length. Error bars displayed are using standard error.

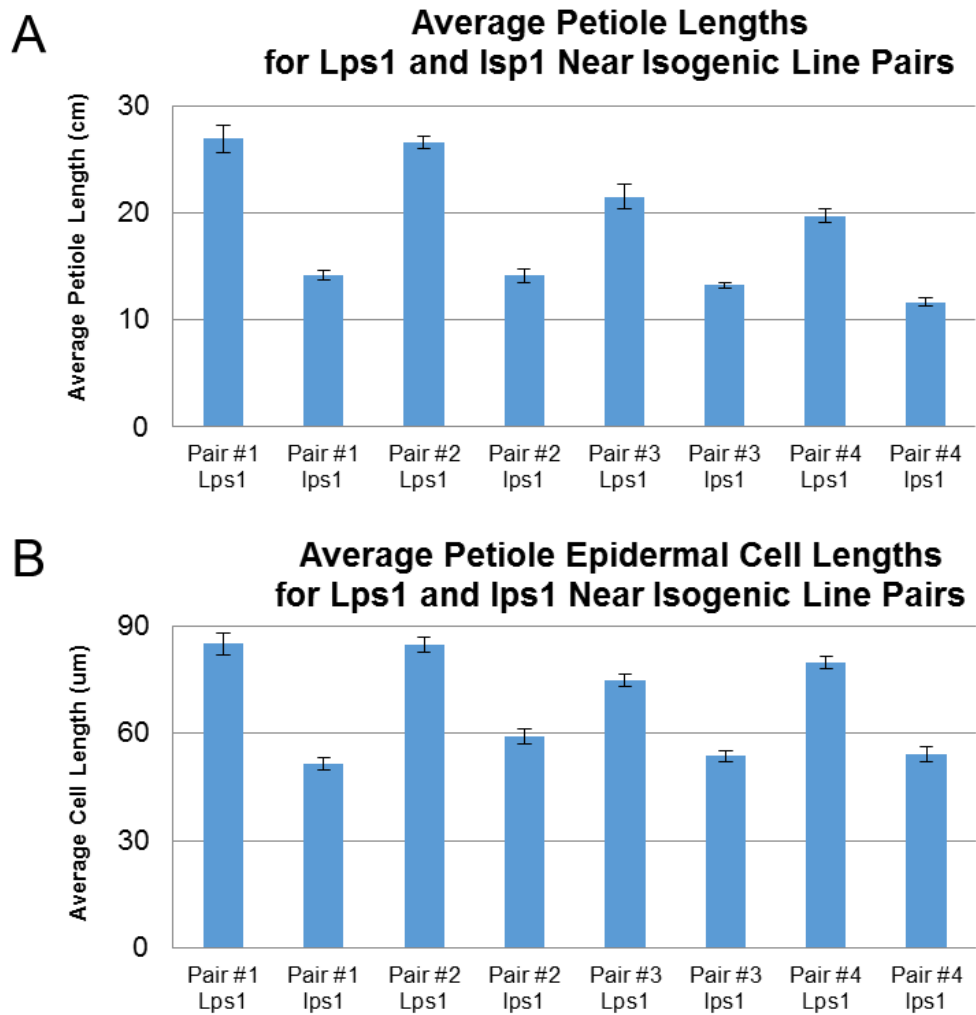
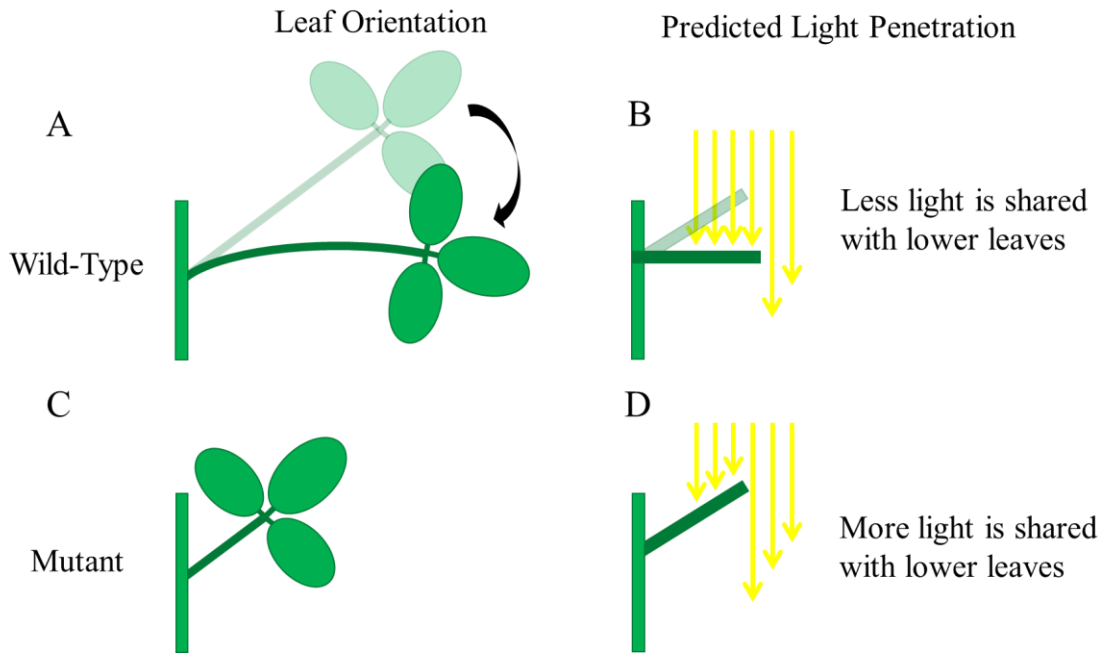


Figure 29. Predicted effects of petiole length on leaf angle of attachment and light penetration into the soybean canopy.

(A) Orientation of a wild-type leaf demonstrating that the longer petiole causes the leaf to have a wider angle of attachment to the stem and is (B) predicted to result in less light penetrating deeper into the soybean canopy. (C) Orientation of a mutant leaf demonstrating that the shorter petiole results in a narrow angle of attachment to the stem and is (D) predicted to result in more light penetrating deeper into the soybean canopy.



Chapter 5 Figures

(No figures)

TABLES

Chapter 1 Tables

Table 1. Soybean mutant populations created using chemical, irradiation, or transposon mutagenesis.

Mutagen: Chemical	Studies
EMS	(Hammond and Fehr, 1983)
EMS & NMU	(Ryan and Harper, 1983)
EMS	(Brossman and Wilcox, 1984)
EMS	(Wilcox et al., 1984)
EMS	(Hammond and Fehr, 1984)
EMS & NaN ₃	(Carroll et al., 1985b)
EthylNitrosourea (ENU)	(Sebastian and Chaleff, 1987)
EMS	(Bubeck et al., 1989)
EMS & NMU	(Gremaud and Harper, 1989)
EMS & NMU	(Sebastian et al., 1989)
NMU	(Fehr et al., 1991)
EMS	(Akao and Kouchi, 1992)
EMS	(Lee et al., 1997)
EMS	(Stijšín et al., 1998)
EMS	(Wilcox et al., 2000)
NMU	(Hitz et al., 2002)
EMS	(Zhu and Sun, 2006)
EMS & NMU	(Cooper et al., 2008)
EMS	(Xia et al., 2012)
Mutagen: Irradiation	Studies
Neutron irradiation	(Humphrey, 1951)
X-ray & Thermal Neutrons	(Rawlings et al., 1958)
γ-ray	(Carroll et al., 1985a)
X-ray	(Takagi et al., 1989)
γ-ray	(Odanaka and Kaizuma, 1989)
X-ray	(Takagi et al., 1990)
γ-ray	(Hajika et al., 1991)
γ-ray	(Kitagawa et al., 1991)

Table 1 (continued)

X-ray	(Rahman et al., 1994)
γ -ray	(Takahashi et al., 1994)
X-ray	(Takagi et al., 1995)
X-ray	(Rahman et al., 1995)
X-ray	(Takagi and Rahman, 1995)
X-ray	(Rahman et al., 1998)
Fast Neutron	(Men et al., 2002)
γ -ray	(Atak et al., 2004)
γ -ray	(Yuan et al., 2007)
X-ray	(Anai et al., 2008)
γ -ray	(Kim et al., 2010)
Ion beam	(Arase et al., 2011)
Fast Neutron	(Bolon et al., 2011)
γ -ray	(Ruddle et al., 2013)
Fast Neutron	(Vincent et al., 2015)
Mutagen: Chemical and Irradiation	Studies
EMS & X-ray	(Hammond and Fehr, 1975)
EMS, NaN ₃ , & γ -ray	(Carroll et al., 1986)
EMS & X-ray	(Watanabe et al., 2009)
EMS & X-ray	(Watanabe et al., 2011)
EMS & X-ray	(Anai, 2012)
EMS & γ -ray	(Kumari et al., 2014)
Fast Neutron & EMS	(Gillman et al., 2014)
EMS & X-ray	(Hoshino et al., 2014)
EMS, DES, & γ -ray	(Gobinath and Pavadai, 2015)
Mutagen: Transposon	Studies
<i>Tgm9</i>	(Palmer et al., 1989, 2008a, 2008b)
<i>Ac/Ds</i>	(Mathieu et al., 2009)
<i>mPing</i>	(Hancock et al., 2011)
<i>Tgm9</i>	(Raval et al., 2013)
<i>Tnt1</i>	(Cui et al., 2013)

Table 2. Basic summary information for different types of transposons used in soybean mutagenesis.

Transposon Population	USDA-APHIS Regulation	Transposition induction by	Transposition can be arrested
<i>Ac/Ds</i>	Transgenic	<i>Ac</i>	Yes
<i>mPing</i>	Transgenic	<i>ORF1 & TPase</i>	Yes
<i>Tnt1</i>	Transgenic	Tissue culture	Yes
<i>Tgm9</i>	Non-transgenic	(Fully autonomous)	No

Chapter 2 Tables

Table 3. PCR Primers used to amplify Glyma13g30560 for *y11* and *y11-2*. Some reactions used a nested primer to amplify a unique sequence, while some reactions used the same set of primers for both reactions (or do not need to use a second reaction to obtain the necessary amplification specificity).

PCR Reaction	Type	Primer Pair First Reaction
1	Forward	TGGCACCCACTAACATTTCC
1	Reverse	CCAGTATCCTTTTTATTTAGGAGACC
2	Forward	TGGCACCCACTAACATTTCC
2	Reverse	CACACAACACACAAAAGAATGG
3	Forward	TGGCACCCACTAACATTTCC
3	Reverse	CACACAACACACAAAAGAATGG
4	Forward	GGCCAGGCCTTTGCATTTTG
4	Reverse	ACTCAGCACACACCTTGGAG
5	Forward	CCTGACTGAGGGTGTCAAGG
5	Reverse	GAAGTTAATCTTGGAGTGTATTTTGC
6	Forward	CCTGACTGAGGGTGTCAAGG
6	Reverse	GAAGTTAATCTTGGAGTGTATTTTGC

PCR Reaction	Type	Primer Pair Second Reaction (Nested)
1	Forward	TGGCACCCACTAACATTTCC
1	Reverse	CCAGTATCCTTTTTATTTAGGAGACC
2	Forward	CGTTTTTGTCTTAAAAGCTTGATT
2	Reverse	GTTTGCAGCACACCATCC
3	Forward	TCTTCTTCTTCCAAGCCTTCC
3	Reverse	CACACAACACACAAAAGAATGG
4	Forward	GGCCAGGCCTTTGCATTTTG
4	Reverse	ACTCAGCACACACCTTGGAG
5	Forward	CCTGACTGAGGGTGTCAAGG
5	Reverse	GGCACTTACGTTGTCTCTTCC
6	Forward	GCTGAGTTGAATGTGGATGG
6	Reverse	GAAGTTAATCTTGGAGTGTATTTTGC

Table 4. PCR Primers used to amplify Glyma15g08680 for CD-5.

Some reactions used a nested primer to amplify a unique sequence, while some reactions used the same set of primers for both reactions (or do not need to use a second reaction to obtain the necessary amplification specificity).

PCR Reaction	Type	Primer Pair First Reaction
1	Forward	CGGAGACTGGTAAATGTGAGC
1	Reverse	CAGCACACCTCCAAAACAAG
2	Forward	GGCTAGGCCTTTGTGTTTGA
2	Reverse	AACGGGAAATGCTGATTGAG
3	Forward	GCTGCTGGATAGGTTTGGAA
3	Reverse	AACGGGAAATGCTGATTGAG
4	Forward	TCAATTGCGGTTCTTCCTTC
4	Reverse	TGGCCTCTGGCTTTCTTTAG
5	Forward	TCAATTGCGGTTCTTCCTTC
5	Reverse	TGGCCTCTGGCTTTCTTTAG

PCR Reaction	Type	Primer Pair Second Reaction (Nested)
1	Forward	CGGAGACTGGTAAATGTGAGC
1	Reverse	CAGCACACCTCCAAAACAAG
2	Forward	GGCTAGGCCTTTGTGTTTGA
2	Reverse	AACGGGAAATGCTGATTGAG
3	Forward	GCTGCTGGATAGGTTTGGAA
3	Reverse	AACGGGAAATGCTGATTGAG
4	Forward	ACGTCAGCTGCTATGAATGG
4	Reverse	GTAGCTCCCAATGGCAAATC
5	Forward	TCAATTGCGGTTCTTCCTTC
5	Reverse	TGGCCTCTGGCTTTCTTTAG

Table 5. Twenty-two genes models present in the fine-mapped interval Gm13: 33,141,206..33,306,556.

Gene	Location	Gene Annotation
Glyma13g30550	Gm13: 33141000..33143927	Transferase family
Glyma13g30560	Gm13: 33144571..33147208	Magnesium chelatase, subunit ChII
Glyma13g30575	Gm13: 33147978..33151645	Peroxisomal NUDIX hydrolase
Glyma13g30590	Gm13: 33154582..33158861	Family not named
Glyma13g30600	Gm13: 33163957..33165210	Zinc finger, C3HC4 type (RING finger)
Glyma13g30610	Gm13: 33171016..33182189	ATP-Dependent RNA Helicase
Glyma13g30620	Gm13: 33183067..33188556	Glutamate-gated kainate-type ion channel receptor subunit GluR5 and related subunits
Glyma13g30625	Gm13: 33193010..33194243	No functional annotation
Glyma13g30630	Gm13: 33193058..33193890	No functional annotation
Glyma13g30650	Gm13: 33194700..33195205	No functional annotation
Glyma13g30640	Gm13: 33194530..33201169	Glutamate-gated kainate-type ion channel receptor subunit GluR5 and related subunits
Glyma13g30660	Gm13: 33204225..33209549	Glutamate-gated kainate-type ion channel receptor subunit GluR5 and related subunits
Glyma13g30670	Gm13: 33210247..33214795	D-Tyr-tRNA (Tyr) deacylase
Glyma13g30680	Gm13: 33221124..33226612	GDSL-like Lipase/Acylhydrolase
Glyma13g30690	Gm13: 33231142..33237718	GDSL-like Lipase/Acylhydrolase
Glyma13g30710	Gm13: 33243653..33244787	Regulation of transcription, DNA-dependent
Glyma13g30720	Gm13: 33247690..33248688	Regulation of transcription, DNA-dependent
Glyma13g30730	Gm13: 33259622..33263250	Pterin carbinolamine dehydratase PCBD/dimerization cofactor of HNF1
Glyma13g30740	Gm13: 33263320..33264787	Protein of unknown function, DUF599
Glyma13g30750	Gm13: 33280224..33286703	Auxin response factor
Glyma13g30760	Gm13: 33291706..33297139	Mlo family
Glyma13g30770	Gm13: 33302261..33303185	Glutaredoxin

Table 6. Soybean CHLI Genes and Expression Data.

Normalized soybean expression data showing that Glyma13g30560 and Glyma15g08680 share similar expression patterns, however Glyma13g30560 showing approximately twice the expression level in leaf tissues (Source: Severin et al., 2010.)

Gene	Young leaf	Flower	One cm pod	Pod shell 10DAF	Pod shell 14DAF	Root	Nodule
Glyma13g30560	106	16	38	41	33	1	1
Glyma15g08680	51	8	16	18	17	1	1
Glyma07g32550	2	3	2	1	1	2	1
Glyma13g24050	0	2	0	0	1	1	1

Gene	Seed 10DAF	Seed 14DAF	Seed 21DAF	Seed 25DAF	Seed 28DAF	Seed 35DAF	Seed 42DAF
Glyma13g30560	2	3	6	8	6	8	3
Glyma15g08680	2	5	7	9	5	7	3
Glyma07g32550	1	1	1	1	0	0	0
Glyma13g24050	1	2	0	1	0	1	0

Chapter 3 Tables

Table 7. Codominant PCR primer triple used to amplify mutant and wild-type alleles

Primer	Primer sequence	Description
B121R	TTTTCACCCCTGTGGTTTTGG	Use Primer B121R with Primer B124R to amplify a 188 bp fragment across the mutant junction of Gm20:2,010,290 and Gm20:2,007,928
B124F	CGAGTGGTACCAGTGGACAT	Use Primer B124F with Primer B124R to amplify a 708 bp fragment in wild-type individuals
B124R	TGTCCCTTTGGCTTTAGT	Use Primer B124R with Primer B121R or Primer B124R to amplify either mutant or wild-type fragments, respectively

Table 8. PCR Primers used to amplification across chromosome rearrangements in Glyma.20G019300

Primer	Primer sequence	Description
B171F	GGTTGGGCATGAAGTGGTTC	Use B171 primer pair to amplify a 605bp fragment depicted in Fig 3F.
B171R	GATGTCCCACGCCGTTAAAG	Use B171 primer pair to amplify a 605bp fragment depicted in Fig 3F.
B172F	GAACCAACAGAGACTCGTGC	Use B172F and B173R primers to amplify a 1441bp fragment across the junction depicted in Fig 3E.
B173R	CCCTGTGGTTTTGGCAAGTT	Use B172F and B173R primers to amplify a 1441bp fragment across the junction depicted in Fig 3E.

Table 9. PCR Primers used to sequence Glyma.20G019300 in line T31 (PI548159)

Primer	Primer sequence	Description
B230F	GCTTCATATTTCTCATTGAAAACC	Amplify 5'UTR section of NAP1, 819bp, sequence with R primer.
B230R	CCACAGAACCAAATTCAAAGC	Amplify 5'UTR section of NAP1, 819bp, sequence with R primer.
B244F	TTGTGGAGTTTGAGAAGCTTAGG	Amplify exons: 1 & 2; sequence with both primers.
B244R	CTGGAATGTGAAAACCTTTGG	Amplify exons: 1 & 2; sequence with both primers.
B231F	CATTCTCCTGCTATCATTGACC	Amplify exons: 3,4,5; sequence with both primers.
B231R	GCTAGATTATACAACATGCTATGTCC	Amplify exons: 3,4,5; sequence with both primers.
B232F	TTTTCCCTAACCATTGTCACC	Amplify exons: 6; sequence with F primer.
B232R	TGCTGCTCCTAAAAGTAGAAAGG	Amplify exons: 6; sequence with F primer.
B233F	TTCATTGTTTTCTAGATTCTTTCC	Amplify exons: 7,8; sequence with both primers.
B233R	TAAACCAACAATTTTCAGTACCC	Amplify exons: 7,8; sequence with both primers.
B234F	CCTTACTAATAGTCATCCAAATGTTGT	Amplify exons: 9,10; sequence with one or both primers.
B234R	CCAAATCACTGAAAATAGCAACC	Amplify exons: 9,10; sequence with one or both primers.
B235F	TGTTGATATTTGTTACTCTTTTCTGG	Amplify exon: 11; sequence with F primer.

Table 9 (continued)

B235R	TGATATGAAACAACAAAAGGAGAGG	Amplify exon: 11; sequence with F primer.
B236F	CTCTTCTGTGGACTCAGTGTGG	Amplify exon: 12; sequence with F primer.
B236R	TCACCCCTTATGTTAGTTTTTGG	Amplify exon: 12; sequence with F primer.
B237F	TGATGAATGGTTTGAAAAATGC	Amplify exon: 13; sequence with F primer.
B237R	GAACGCATCTATTTGCATGG	Amplify exon: 13; sequence with F primer.
B238F	TTTTGCATGGGTGTTTTGG	Amplify exons: 14, 15; sequence with both primers.
B238R	GATCATGATTTTGA CTATAACCATCG	Amplify exons: 14, 15; sequence with both primers.
B239F	TGTAACAGCTGAGTTAGAGCTTCC	Amplify exons: 16, 17; sequence with both primers.
B239R	ATGCCTCCATCAAAATGTGC	Amplify exons: 16, 17; sequence with both primers.
B240F	TGTGTGATGGTAGCAATATGTGG	Amplify exons: 18, 19, 20; sequence with both primers.
B240R	TTTCTTTAAAGGGCGATACCC	Amplify exons: 18, 19, 20; sequence with both primers.
B241F1	TCCAGATCCAACATTAGTCACC	Amplify exon: 21; Amplify with F1, R. Sequence with all 3 primers.
B241F2	AGGACCGGTTTCTTCTCTGC	Amplify exon: 21; Amplify with F1, R. Sequence with all 3 primers.
B241R	CAATGGCAATGAATAGTTCAGC	Amplify exon: 21; Amplify with F1, R. Sequence with all 3 primers.
B242F	GCTGAACTATTCATTGCCATTG	Amplify exon: 22; sequence with F primers.
B242R	AATATATTGCAACATTGCCTACC	Amplify exon: 22; sequence with F primers.
B243F1	GAATTCACCTGCGGTTTTG	Amplify exon: 23 and 3'UTR; Amplify with F1, R. sequence with all 3 primers.
B243F2	TATTATGGGCACGAATCAGG	Amplify exon: 23 and 3'UTR; Amplify with F1, R. Sequence with all 3 primers.
B243R	ACTTGTACTCGAGCGGCATT	Amplify exon: 23 and 3'UTR; Amplify with F1, R. Sequence with all 3 primers.

Table 10. PCR Primers used to test for the presence of the *GmNAPI* construct in Arabidopsis T₁ individuals

Primer	Primer sequence	Description
B182F	ATTCGTGCTTACAACCTCGCC	Use primers B182F and B182R to amplify a 548 bp band from Soybean plants or a 556 bp band from Arabidopsis plants with the <i>GmNAPI</i> construct.
B182R	CCTGAGACTGCCCATCATGA	Use primers B182F and B182R to amplify a 548 bp band from Soybean plants or a 556 bp band from Arabidopsis plants with the <i>GmNAPI</i> construct.

Chapter 4 Tables

Table 11. Primers and enzymes used to test candidate *lsp1* polymorphisms.

	Location	Polymorphism	Polymorphism Location
1	Gm16:36848410	SNP (C/T) (Synonymous)	Glyma.16g209100
2	Gm16:36849837	INDEL (T/TAAG)	Glyma.16g209100
3	Gm16:36860276	Deletion (130 bp)	Glyma.16g209300
4	Gm16:36865819	Deletion (244 bp)	Between Glyma.16g209300 and Glyma.16g209400
5	Gm16:36869783	Deletion (51 bp)	Between Glyma.16g209300 and Glyma.16g209400
6	Gm16:36882978	SNP (A/T)	Between Glyma.16g209400 and Glyma.16g209500
7	Gm16:36883605	Deletion (47 bp)	Between Glyma.16g209400 and Glyma.16g209500
8	Gm16:36898863	SNP (C/T)	Glyma.16g209600

	CAPS Enzyme	Name	Sequence
1	BccI	B185F	GCATAGCTAGCATATCCATCAGGG
		B185R	TCCTTGGTTAGCTTCACACC
2	EarI	B194F	CGAAAAGAAGGGCATATTTTACC
		B194R	ATCATAAGCCGAGCAAAAACG
3	na	B187F	ACCTGCAGAAAGCCTTTGG
		B187R	AGTCACAACCCTGAACATCG
4	na	B188F	CATGTCAGCCCTATAGTATCTGC
		B188R	CAAAATCAATGTTGAAAGTCTTACC
5	na	B189F	CGTAGAAGCACAAAGTGTCTCC
		B189R	CCACTTGCGATCAAATTAAGC
6	AseI	B190F	TTTCATTCTCTCTAAGTGTCAAAAGG
		B190R	TCATAACATCCATTATGATTTCTCG
7	na	B191F	GCCTGCACAACACTTGGTTA
		B191R	ATCATGCGCCCACTGTAAA
8	BccI	B184F	TGGAGCAGTGCTACGATTTG
		B184R	GGCAAACCACTCATGCTTTT

Table 12. D76-1609 and its parental haplotypes in and adjacent to the *lpsI* fine mapped interval.

SoySNP50K genotype data are provide for D75-1609, Forrest, PI 229358, and Lee. PI 95960 did not have SoySNP50K data. Only polymorphic SoySNP50K positions are shown. The genotypes of the deletion variants are given by a number assigned to the allele type. Important genotype scores are bolded. The candidate *lpsI* INDEL at Polymorphism #2 is shaded. The SoySNP50K positions that match D76-1609 are colored in grey. The haplotypes of the parents suggest that Forrest donated the haplotype containing the fine mapped interval to D76-1609 (with the exception of the AAG insertion). Positions are given according to the version 2 genome assembly, Glyma.Wm82.a2.v1.

Chr	Position	Variant	D76-1609 PI548256	Forrest PI548655	Soden- daizu PI229358	Lee PI548656	PI 95960	Type
Gm16	36767455	ss715624897	C	T	C	C		SoySNP50K
Gm16	36774600	ss715624899	C	T	C	C		SoySNP50K
Gm16	36781107	ss715624901	G	A	G	G		SoySNP50K
Gm16	36782323	ss715624902	A	A	A	G		SoySNP50K
Gm16	36797859	ss715624904	G	G	G	A		SoySNP50K
Gm16	36808778	ss715624905	A	A	A	G		SoySNP50K
<i>Edge of the fine mapping interval</i>								
Gm16	36831753	ss715624906	G	G	G	A		SoySNP50K
Chr16	36840227	ss715624908	C	C	A	A		SoySNP50K
Gm16	36848410	Polymorphism #1	T	T	T	C	C	SNP
Gm16	36849837	Polymorphism #2	TAAG	T	T	T	T	Candidate
Gm16	36860276	Polymorphism #3	1	1	1	2	2	Deletion
Gm16	36865819	Polymorphism #4	1	1	1	2	2	Deletion
Gm16	36869783	Polymorphism #5	1	1	1	2	2	Deletion
Gm16	36882978	Polymorphism #6	T	T	?	A	T	SNP
Gm16	36883605	Polymorphism #7	1	1	1	3(?)	2	Deletion
Gm16	36898863	Polymorphism #8	T	T	C	C	C	SNP
Gm16	36916624	ss715624913	A	A	A	C		SoySNP50K
<i>Edge of the fine mapping interval</i>								
Gm16	36919622	ss715624914	G	G	G	T		SoySNP50K
Gm16	36927970	ss715624915	T	G	G	T		SoySNP50K
Gm16	36949935	ss715624918	T	T	T	C		SoySNP50K
Gm16	36986699	ss715624921	G	G	A	G		SoySNP50K
Gm16	36990907	ss715624923	A	A	G	A		SoySNP50K

Table 13. NIL average vegetative and reproductive growth stages and average plant height.

Growth stages are averaged across the NIL pairs at the St. Paul 2015 yield trail.

Vegetative growth stages are given up to just past V9 and reproductive stages are given up through R6. Plant heights and growth stages were taken on the same day, with the exception of the first plant height measurement date which occurred one day before the growth stage date.

Date	Days after planting	Ave. V stage of <i>lps</i> NILs	Ave. V stage of <i>Lps1</i> NILs	Ave. R stage of <i>lps1</i> NILs	Ave. R stage of <i>Lps1</i> NILs	Growth stage p-value:	Ave. Height of <i>lps1</i> NILs (cm)	Ave. Height of <i>Lps1</i> NILs (cm)	Plant height p-value:
6.25.2015	24	1.80	1.95			0.055	15.7	15.7	0.9714
6.30.2015	29	2.85	2.98			0.096			
7.01.2015	31	3.48	3.53			0.428	21.4	22.3	0.1691
7.07.2015	36	4.63	4.73			0.258	30.1	31.3	0.1699
7.09.2015	39	6.47	6.61			0.289			
7.14.2015	43	7.03	7.15			0.310	44.1	47.2	0.0267
7.17.2015	46	7.58	7.80			0.095			
7.21.2015	50	8.21	8.29			0.596	54.5	59.3	0.0128
7.24.2015	53	9.58	9.85			0.096			
7.29.2015	58			1.89	1.88	1.000	75.6	83.3	0.0000
8.04.2015	63			2.10	2.20	0.428	86.0	92.0	0.0049
8.11.2015	70			2.85	2.90	0.716	97.2	104.7	0.0002
8.18.2015	77			4.30	4.25	0.772			
8.26.2015	85			5.10	5.10	1.000			
9.08.2015	99			5.98	6.00	0.330			

Table 14. Preliminary yield trial results from *Lps1* and *lps1* Near Isogenic Line yield trials.

Plot	Loc	Line	lodging score	grain yield mass (g)/0.76m ²	grain yield mass (g)/0.76m ² adjusted 13% moisture	Total mass stems + seeds (g)	Harvest index	Moisture	Protein in Dry basis	Oil Dry basis
1	STP	981	1	271.5	288.6	700	0.4123	6.7	38.19	21.37
2	STP	983	1	232	246.1	600	0.4101	6.93	40.22	22.23
3	STP	984	1	219.4	232.9	560	0.4159	6.84	39.21	21.39
4	STP	988	1	206.1	219.9	560	0.3927	6.3	41.13	21.73
5	STP	Heinong51	3	297.4	317.1	800	0.3963	6.39	39.16	21.04
6	STP	MN1410	1	279.9	299.3	720	0.4157	6.07	40.00	22.76
7	STP	Ozzie	1.75	247.7	260.7	630	0.4138	7.76	40.52	21.71
8	STP	kengfeng16	1.75	242.7	258.0	590	0.4373	6.7	39.20	20.86
9	STP	989	1	208.7	220.3	530	0.4156	7.46	41.00	20.59
10	STP	985	2.5	149.7	159.3	830	0.1919	6.59	41.10	19.87
11	STP	982	1	500.7	531.8	1140	0.4665	6.78	39.47	23.01
12	STP	980	2	212.2	224.6	560	0.4011	7.16	39.62	20.71
13	STP	MN1410	1.5	364.1	388.1	890	0.4361	6.41	40.22	22.89
14	STP	980	2.5	289.6	308.2	710	0.4341	6.57	37.63	21.91
15	STP	MN1410	3	286.5	305.0	720	0.4236	6.54	40.17	23.11
16	STP	984	2	243.2	259.2	630	0.4114	6.43	38.04	21.97
17	STP	MN1410	2.5	264.7	279.5	680	0.4111	7.4	41.38	23.14
18	STP	984	2.5	340.5	362.7	850	0.4267	6.49	39.47	20.37
19	STP	985	2	245	260.4	680	0.3830	6.71	41.00	20.94
20	STP	993	3.5	353.6	373.9	910	0.4109	7.26	45.17	19.71
21	STP	985	1.5	217.8	230.1	680	0.3384	7.35	41.81	20.64
22	STP	982	1.5	239.6	254.5	600	0.4242	6.77	40.71	22.27
23	STP	981	3	319	340.4	840	0.4052	6.29	38.89	20.94
24	STP	978	1.5	312.5	335.0	730	0.4589	5.8	37.91	20.81
25	STP	979	1.5	283.7	301.5	750	0.4019	6.74	38.07	21.39
26	STP	MN1410	3	320.9	340.6	780	0.4367	6.86	41.52	22.77
27	STP	983	2.5	318.1	338.1	870	0.3887	6.7	41.73	22.04
28	STP	MN1410	3.5	379.6	404.1	920	0.4392	6.55	40.75	22.94
29	STP	992	3	292.7	311.5	740	0.4209	6.59	42.38	20.18
30	STP	MN1410	3	384	407.7	1000	0.4077	6.82	39.64	22.91
31	STP	995	4.5	254.3	268.3	720	0.3726	7.51	44.76	19.61

Table 14 continued

32	STP	991	<u>4.5</u>	263	280.1	700	0.4002	6.48	43.68	18.59
33	STP	986	3	203.4	214.4	560	0.3829	7.59	40.60	19.98
34	STP	988	1.5	230	243.5	690	0.3529	7.12	41.73	20.98
35	STP	994	4	195.9	207.8	550	0.3778	6.94	43.76	19.57
36	STP	990	4	268.1	282.1	650	0.4340	7.79	45.50	18.52
37	STP	991	4	180	189.7	690	0.2749	7.63	44.63	19.34
38	STP	996	4	147.7	155.1	580	0.2674	7.98	41.67	21.21
39	STP	989	1.5	240.3	254.0	660	0.3849	7.28	40.90	20.60
40	STP	987	3	314.6	333.1	820	0.4062	7.13	40.39	20.75
41	STP	990	4	280.4	295.9	730	0.4053	7.49	43.40	19.12
42	STP	994	3.5	262.8	278.6	660	0.4222	6.97	43.57	19.51
43	STP	997	4	107.6	114.1	580	0.1967	6.99	42.52	20.54
44	STP	MN1410	3.5	328.2	350.4	770	0.4550	6.24	40.66	22.26
45	STP	1000	3	293.9	312.9	790	0.3960	6.55	43.67	19.92
46	STP	MN1410	3.5	261.1	278.6	670	0.4159	6.29	41.31	22.27
47	STP	995	4	210.3	221.6	640	0.3463	7.62	45.11	18.22
48	STP	MN1410	3.5	366	385.8	860	0.4486	7.58	40.63	22.71
49	STP	992	4	250.6	264.7	740	0.3576	7.39	46.89	17.97
50	STP	997	4.75	128	136.6	520	0.2626	6.32	40.16	21.16
51	STP	998	4	201.8	212.7	650	0.3273	7.59	42.88	20.25
52	STP	1001	3.5	198	210.2	550	0.3821	6.86	43.10	20.12
53	STP	999	4	163.5	173.0	560	0.3089	7.21	40.95	20.68
54	STP	Norchief	5	283.3	301.4	710	0.4245	6.61	40.74	22.77
55	STP	MN1410	3.5	371.8	396.5	870	0.4558	6.35	41.25	20.94
56	STP	998	3.5	202.6	214.6	540	0.3974	7.08	41.98	20.65
57	STP	MN1410	4	280.4	297.0	700	0.4242	7.09	40.29	22.85
58	STP	999	4.5	114.6	121.4	370	0.3281	7.06	41.86	20.99
59	STP	MN1410	3.75	254.3	267.6	630	0.4248	7.76	41.12	22.09
60	STP	993	4	142.3	149.6	460	0.3252	7.87	45.80	17.77
61	RO	MN1410	3	276.3	294.1	852.753	0.3449	6.55	41.95	20.73
62	RO	998	4	146.7	155.2	535.239	0.2900	7.19	40.75	20.58
63	RO	MN1410	2	213.4	229.7	589.67	0.3895	5.36	43.3	20.55
64	RO	990	4	182.1	190.9	798.322	0.2392	8.15	44.71	17.91
65	RO	MN1410	2	257.2	273.0	698.532	0.3909	6.84	44.60	20.71
66	RO	982	2	289.2	306.6	816.466	0.3755	7	42.65	21.03
67	RO	MN1410	3	251.4	267.2	653.172	0.4091	6.7	42.76	20.36
68	RO	990	3.75	180.2	190.3	940	0.2025	7.37	44.17	18.95

Table 14 continued

69	RO	991	4	227.7	241.0	980	0.2459	7.18	44.87	18.49
70	RO	Ozzie	2	231.5	247.9	570	0.4349	5.93	40.48	21.60
71	RO	983	2	228.8	243.7	640	0.3808	6.48	40.51	21.81
72	RO	1000	1.5	177.6	187.0	570	0.3281	7.7	44.04	19.45
73	RO	991	3	177.9	188.5	750	0.2513	7.05	43.75	18.59
74	RO	981	3	181.2	191.9	640	0.2998	7.11	40.04	19.07
75	RO	999	3	161.8	171.6	670	0.2561	6.96	41.13	19.93
76	RO	994	3.5	154.1	162.2	600	0.2704	7.72	45.86	17.28
77	RO	995	4.5	201.3	212.1	990	0.2143	7.62	45.68	17.85
78	RO	MN1410	2.5	270.9	287.1	740	0.3880	7.01	41.48	22.30
79	RO	980	2.5	191.4	202.1	620	0.3260	7.4	38.92	20.20
80	RO	MN1410	2.5	174.8	186.9	490	0.3815	6.07	42.40	20.51
81	RO	1001	2.5	168.3	177.1	570	0.3108	7.75	44.77	18.54
82	RO	MN1410	2	211.8	223.8	560	0.3997	7.33	43.37	20.81
83	RO	Norchief	5	124	132.2	380	0.3478	6.42	41.24	21.92
84	RO	MN1410	2.5	193	205.0	560	0.3660	6.79	44.11	20.65
85	RO	980	3	164	173.5	680	0.2552	7.2	40.13	19.16
86	RO	Heinong51	3.5	233.3	247.1	800	0.3089	7.08	42.61	18.47
87	RO	993	2	137.2	143.6	470	0.3056	8.32	46.57	18.18
88	RO	978	1.25	172.7	182.5	620	0.2943	7.33	39.73	19.14
89	RO	994	1.5	199.8	212.6	690	0.3082	6.58	45.24	18.58
90	RO	988	1.25	185	197.8	550	0.3596	6.1	40.02	20.83
91	RO	989	1	204.7	218.3	690	0.3163	6.38	40.01	21.16
92	RO	997	4	95.7	100.9	660	0.1528	7.59	40.09	21.34
93	RO	Heinong51	3.5	119.8	127.3	490	0.2599	6.7	41.28	19.08
94	RO	kengfeng16	3.5	171.4	181.0	470	0.3851	7.39	40.13	20.76
95	RO	989	1	209.1	223.9	680	0.3292	5.93	40.07	20.09
96	RO	995	2	130.3	137.2	670	0.2048	7.7	46.10	18.21
97	RO	979	1.5	195.5	206.7	610	0.3389	7.25	41.84	18.61
98	RO	992	2.5	161.9	171.5	480	0.3573	7.06	45.07	17.85
99	RO	kengfeng16	4	192	202.9	510	0.3979	7.3	41.67	20.30
100	RO	981	2.5	166.1	176.2	600	0.2937	6.92	40.81	18.37
101	RO	998	2	114	120.7	410	0.2945	7.08	41.62	20.82
102	RO	MN1410	2	171.7	181.3	460	0.3942	7.39	42.89	21.18
103	RO	987	1.5	197.7	209.2	670	0.3123	7.16	40.21	19.22
104	RO	MN1410	2.5	98.1	103.8	580	0.1789	7.21	42.96	21.15
105	RO	996	3.5	279.4	296.7	560	0.5298	6.82	43.17	21.14

Table 14 continued

106	RO	MN1410	2	228.9	243.6	640	0.3807	6.57	43.42	20.64
107	RO	984	2.5	188	199.8	620	0.3222	6.73	39.55	19.53
108	RO	MN1410	3	217.3	231.3	590	0.3921	6.55	41.00	22.49
109	RO	983	2.5	139.9	147.9	560	0.2642	7.26	40.09	21.73
110	RO	985	1.5	182.7	194.4	650	0.2991	6.58	41.38	19.47
111	RO	992	2.5	158.9	167.8	520	0.3226	7.43	45.77	18.25
112	RO	997	3	99.5	105.7	550	0.1922	6.76	41.04	20.49
113	RO	985	1.5	154.1	163.5	560	0.2919	6.93	40.83	19.52
114	RO	986	1.25	127.3	134.8	430	0.3135	7.09	41.76	18.28
115	RO	985	1.5	133.7	141.9	460	0.3084	6.9	41.59	19.26
116	RO	999	2	140.3	148.5	550	0.2701	7.13	43.14	18.88
117	RO	MN1410	2.5	236.2	252.4	590	0.4277	6.16	41.98	21.53
118	RO	984	2	162.4	173.1	490	0.3532	6.43	40.15	20.2
119	RO	MN1410	2	164.9	175.0	480	0.3646	6.88	41.25	21.64
120	RO	Green MinnGold	1							
121	RO	MN1410	1.5	159.7	170.8	450	0.3797	6.02	41.79	21.41
122	RO	993	2	143.5	151.0	480	0.3145	7.79	46.32	17.91
123	RO	MN1410	2.5	193.3	206.3	530	0.3892	6.29	42.28	21.51
124	RO	982	1.5	188.1	199.9	590	0.3389	6.71	38.82	22.06

*note RO had wet plants

Table 15. Additional data from preliminary yield trial results from *Lps1* and *lps1* Near Isogenic Line yield trials.

Plo t	Loc	Line	Pheno	Plant # in sample area	Line	Full line name	Outcross parent	Type	Notes
1	STP	981	WT	23	BWC 3073-08	M12-601A083-04	Heinong51	WT	
2	STP	983	WT	26	BWC 3109-08	M12-601A144-09	Heinong51	WT	
3	STP	984	MT	18	BWC 3123-02	M12-602A009-01	Kengfeng16	MT	
4	STP	988	MT	23	BWC 3144-01	M12-602A026-01	Kengfeng16	MT	
5	STP	Heinong51		31					
6	STP	MN1410		24					
7	STP	Ozzie		25					1.5 to 2
8	STP	kengfeng16		28					1.5 to 2
9	STP	989	WT	22	BWC 3144-05	M12-602A026-01	Kengfeng16	WT	
10	STP	985	WT	25	BWC 3123-12	M12-602A009-01	Kengfeng16	WT	
11	STP	982	MT	31	BWC 3109-01	M12-601A144-09	Heinong51	MT	
12	STP	980	MT	18	BWC 3073-03	M12-601A083-04	Heinong51	MT	1 to 3
13	STP	MN1410		33					
14	STP	980	MT	19	BWC 3073-03	M12-601A083-04	Heinong51	MT	2 to 3
15	STP	MN1410		30					
16	STP	984	MT	16	BWC 3123-02	M12-602A009-01	Kengfeng16	MT	1 to 3
17	STP	MN1410		23					2 to 3
18	STP	984	MT	18	BWC 3123-02	M12-602A009-01	Kengfeng16	MT	2 to 3
19	STP	985	WT	20	BWC 3123-12	M12-602A009-01	Kengfeng16	WT	
20	STP	993	WT	27	BWC 3228-11	M12-603019-03	MN1410	WT	3 to 4
21	STP	985	WT	17	BWC 3123-12	M12-602A009-01	Kengfeng16	WT	
22	STP	982	MT	19	BWC 3109-01	M12-601A144-09	Heinong51	MT	
23	STP	981	WT	28	BWC 3073-08	M12-601A083-04	Heinong51	WT	
24	STP	978	MT	25	BWC 3051-01	M12-601A073-01	Heinong51	MT	
25	STP	979	WT	32	BWC 3051-11	M12-601A073-01	Heinong51	WT	
26	STP	MN1410		29					
27	STP	983	WT	30	BWC 3109-08	M12-601A144-09	Heinong51	WT	2 to 3
28	STP	MN1410		33					
29	STP	992	MT	28	BWC 3228-04	M12-603019-03	MN1410	MT	
30	STP	MN1410		36					
31	STP	995	WT	17	BWC 3238-05	M12-603022-01	MN1410	WT	4 to 5
32	STP	991	WT	19	BWC 3219-06	M12-603008-06	MN1410	WT	4 to 5
33	STP	986	MT	12	BWC 3130-01	M12-602A009-08	Kengfeng16	MT	
34	STP	988	MT	26	BWC 3144-01	M12-602A026-01	Kengfeng16	MT	1 to 2
35	STP	994	MT	18	BWC 3238-01	M12-603022-01	MN1410	MT	
36	STP	990	MT	20	BWC 3219-02	M12-603008-06	MN1410	MT	
37	STP	991	WT	20	BWC 3219-06	M12-603008-06	MN1410	WT	
38	STP	996	MT	20	BWC 3238-02	M12-603022-01	MN1410	MT	
39	STP	989	WT	24	BWC 3144-05	M12-602A026-01	Kengfeng16	WT	1 to 2
40	STP	987	WT	15	BWC 3130-07	M12-602A009-08	Kengfeng16	WT	
41	STP	990	MT	22	BWC 3219-02	M12-603008-06	MN1410	MT	
42	STP	994	MT	18	BWC 3238-01	M12-603022-01	MN1410	MT	
43	STP	997	WT	22	BWC 3238-10	M12-603022-01	MN1410	WT	between 2 and 5, mostly 4 to 5
44	STP	MN1410		30					

Table 15 Continued

45	STP	1000	MT	19	BWC 3350-02	M12-609012-03	Ozzie	MT	
46	STP	MN1410		33					
47	STP	995	WT	19	BWC 3238-05	M12-603022-01	MN1410	WT	
48	STP	MN1410		31					2 to 4 to 5, mostly 4
49	STP	992	MT	21	BWC 3228-04	M12-603019-03	MN1410	MT	
50	STP	997	WT	14	BWC 3238-10	M12-603022-01	MN1410	WT	4.5 to 5
51	STP	998	MT	21	BWC 3328-01	M12-608038-01	Norchief	MT	
52	STP	1001	WT	11	BWC 3350-05	M12-609012-03	Ozzie	WT	
53	STP	999	WT	12	BWC 3328-09	M12-608038-01	Norchief	WT	
54	STP	Norchief		23					
55	STP	MN1410		29					2 to 4
56	STP	998	MT	18	BWC 3328-01	M12-608038-01	Norchief	MT	
57	STP	MN1410		27					
58	STP	999	WT	13	BWC 3328-09	M12-608038-01	Norchief	WT	4 to 5
59	STP	MN1410		21					3.5 to 4
60	STP	993	WT	18	BWC 3228-11	M12-603019-03	MN1410	WT	
61	RO	MN1410		44					good stand
62	RO	998	MT	22	BWC 3328-01	M12-608038-01	Norchief	MT	gophers digging up plot
63	RO	MN1410		25					good stand, gophers digging up plot
64	RO	990	MT	30	BWC 3219-02	M12-603008-06	MN1410	MT	late maturing
65	RO	MN1410		32					good stand
66	RO	982	MT	32	BWC 3109-01	M12-601A144-09	Heinong51	MT	good stand
67	RO	MN1410		28					good stand
68	RO	990	MT	22	BWC 3219-02	M12-603008-06	MN1410	MT	later maturing
69	RO	991	WT	23	BWC 3219-06	M12-603008-06	MN1410	WT	later maturing
70	RO	Ozzie		25					thinner stand
71	RO	983	WT	36	BWC 3109-08	M12-601A144-09	Heinong51	WT	good stand
72	RO	1000	MT	23	BWC 3350-02	M12-609012-03	Ozzie	MT	later maturing, short plants, good stand
73	RO	991	WT	28	BWC 3219-06	M12-603008-06	MN1410	WT	later maturing
74	RO	981	WT	25	BWC 3073-08	M12-601A083-04	Heinong51	WT	ok stand
75	RO	999	WT	18	BWC 3328-09	M12-608038-01	Norchief	WT	ok stand
76	RO	994	MT	23	BWC 3238-01	M12-603022-01	MN1410	MT	later maturing
77	RO	995	WT	26	BWC 3238-05	M12-603022-01	MN1410	WT	later maturing
78	RO	MN1410		38					ok stand
79	RO	980	MT	20	BWC 3073-03	M12-601A083-04	Heinong51	MT	more space stand
80	RO	MN1410		31					more space stand
81	RO	1001	WT	21	BWC 3350-05	M12-609012-03	Ozzie	WT	ok slightly later
82	RO	MN1410		35					ok stand
83	RO	Norchief		30					poorer stand
84	RO	MN1410		35					ok stand
85	RO	980	MT	27	BWC 3073-03	M12-601A083-04	Heinong51	MT	ok stand, late maturing
86	RO	Heinong51		35					ok stand
87	RO	993	WT	32	BWC 3228-11	M12-603019-03	MN1410	WT	good stand
88	RO	978	MT	36	BWC 3051-01	M12-601A073-01	Heinong51	MT	short plants, very upright
89	RO	994	MT	36	BWC 3238-01	M12-603022-01	MN1410	MT	good stand

Table 15 Continued

90	RO	988	MT	29	BWC 3144-01	M12-602A026-01	Kengfeng16	MT	ok stand, Low on seed, pop at 88% 42.9g not full 48.6g
91	RO	989	WT	30	BWC 3144-05	M12-602A026-01	Kengfeng16	WT	good stand, b/c 988 pair plot low on seed, decreased seed to 46.6g from 52.8
92	RO	997	WT	21	BWC 3238-10	M12-603022-01	MN1410	WT	late maturing
93	RO	Heinong51		34					ok stand, but gopher holes
94	RO	Kengfeng16		34					ok stand some plants 1 other 5 lodging
95	RO	989	WT	35	BWC 3144-05	M12-602A026-01	Kengfeng16	WT	good stand
96	RO	995	WT	27	BWC 3238-05	M12-603022-01	MN1410	WT	later maturity
97	RO	979	WT	33	BWC 3051-11	M12-601A073-01	Heinong51	WT	good stand
98	RO	992	MT	34	BWC 3228-04	M12-603019-03	MN1410	MT	good stand
99	RO	kengfeng16		36					some lodge 2-5 ok stand
100	RO	981	WT	24	BWC 3073-08	M12-601A083-04	Heinong51	WT	lodge 2-3 ok stand
101	RO	998	MT	24	BWC 3328-01	M12-608038-01	Norchief	MT	uneven/poor stand, seg determinate and indeterminate?
102	RO	MN1410		37					ok stand
103	RO	987	WT	22	BWC 3130-07	M12-602A009-08	Kengfeng16	WT	ok stand
104	RO	MN1410		38					ok stand
105	RO	996	MT	19	BWC 3238-02	M12-603022-01	MN1410	MT	poor stand-- weeds in plot previously
106	RO	MN1410		37					ok stand
107	RO	984	MT	24	BWC 3123-02	M12-602A009-01	Kengfeng16	MT	ok stand
108	RO	MN1410		27					gophers digging up plot-- poorer stand
109	RO	983	WT	28	BWC 3109-08	M12-601A144-09	Heinong51	WT	gophers digging up plot-- poorer stand
110	RO	985	WT	24	BWC 3123-12	M12-602A009-01	Kengfeng16	WT	ok stand
111	RO	992	MT	33	BWC 3228-04	M12-603019-03	MN1410	MT	ok stand
112	RO	997	WT	31	BWC 3238-10	M12-603022-01	MN1410	WT	poor stand
113	RO	985	WT	21	BWC 3123-12	M12-602A009-01	Kengfeng16	WT	ok stand
114	RO	986	MT	32	BWC 3130-01	M12-602A009-08	Kengfeng16	MT	ok stand
115	RO	985	WT	29	BWC 3123-12	M12-602A009-01	Kengfeng16	WT	ok stand
116	RO	999	WT	35	BWC 3328-09	M12-608038-01	Norchief	WT	ok stand
117	RO	MN1410		47					ok stand
118	RO	984	MT	30	BWC 3123-02	M12-602A009-01	Kengfeng16	MT	ok stand, some weeds?
119	RO	MN1410		39					ok stand
120	RO	Green MinnGold		24					green minngold
121	RO	MN1410		38					ok stand
122	RO	993	WT	30	BWC 3228-11	M12-603019-03	MN1410	WT	ok stand
123	RO	MN1410		34					ok stand
124	RO	982	MT	24	BWC 3109-01	M12-601A144-09	Heinong51	MT	ok stand, gophers digging up plot

Table 16. Examples of comparisons made between *Lps1* and *lps1* NIL pairs. The following table lists the comparisons between the *Lps1* and *lps1* NIL pairs for the 2015 preliminary yield trial. Not all pairs of plots were compared if there was reason to suspect that factors outside of the genetic makeup influenced the observed yields. For example, select pairs of plots at the Rosemount location were dropped from the comparison if one or both was damaged by pocket gophers. Comparisons were done within each NIL pair relative to the yield of the wild-type *Lps1* NIL.

Mutant values as a percent of WT values						(Mutant - WT) / WT						
Mutant Plot	total mass adjusted for % moisture	Harvest index	Protein Dry basis	Oil Dry basis	Biomass	WT Plot	All WT values expressed as 100% of WT value for all traits	total mass adjusted for % moisture	Harvest index	Protein Dry basis	Oil Dry basis	biomass
3	146.2	216.7	95.4	107.6	67.5	10	100	0.462	1.167	-0.046	0.076	-0.325
4	99.8	94.5	100.3	105.5	105.7	9	100	-0.002	-0.055	0.003	0.055	0.057
12	77.8	97.3	103.7	96.9	80.0	1	100	-0.222	-0.027	0.037	0.031	-0.200
14	90.5	107.1	96.8	104.6	84.5	23	100	-0.095	0.071	-0.032	0.046	-0.155
16	112.6	121.6	91.0	106.4	92.6	21	100	0.126	0.216	-0.090	0.064	-0.074
18	139.3	111.4	96.3	97.3	125.0	19	100	0.393	0.114	-0.037	0.027	0.250
22	75.3	109.1	97.6	101.0	69.0	27	100	-0.247	0.091	-0.024	0.010	-0.310
24	111.1	114.2	99.6	97.3	97.3	25	100	0.111	0.142	-0.004	0.027	-0.027
29	83.3	102.4	93.8	102.4	81.3	20	100	-0.167	0.024	-0.062	0.024	-0.187
33	64.4	94.3	100.5	96.3	68.3	40	100	-0.356	-0.057	0.005	0.037	-0.317
34	95.9	91.7	102.0	101.8	104.5	39	100	-0.041	-0.083	0.020	0.018	0.045
35	93.8	109.1	97.0	107.4	85.9	47	100	-0.062	0.091	-0.030	0.074	-0.141
36	148.7	157.9	101.9	95.8	94.2	37	100	0.487	0.579	0.019	0.042	-0.058
38	113.6	101.8	103.8	100.2	111.5	50	100	0.136	0.018	0.038	0.002	0.115
41	105.6	101.3	99.4	102.9	104.3	32	100	0.056	0.013	-0.006	0.029	0.043
42	103.9	113.3	97.3	99.5	91.7	31	100	0.039	0.133	-0.027	0.005	-0.083
45	148.9	103.6	101.3	99.0	143.6	52	100	0.489	0.036	0.013	0.010	0.436
49	176.9	110.0	102.4	101.1	160.9	60	100	0.769	0.100	0.024	0.011	0.609
51	175.2	99.7	102.4	96.5	175.7	58	100	0.752	-0.003	0.024	0.035	0.757
56	124.1	128.7	102.5	99.9	96.4	53	100	0.241	0.287	0.025	0.001	-0.036
64	101.3	95.2	102.2	96.3	106.4	73	100	0.013	-0.048	0.022	0.037	0.064
66	125.8	98.6	105.3	96.4	127.6	71	100	0.258	-0.014	0.053	0.036	0.276
68	79.0	82.4	98.4	102.5	95.9	69	100	-0.210	-0.176	-0.016	0.025	-0.041
72	105.6	105.6	98.4	104.9	100.0	81	100	0.056	0.056	-0.016	0.049	0.000
76	76.5	126.2	100.4	96.8	60.6	77	100	-0.235	0.262	0.004	0.032	-0.394

Table 16 Continued

79	105.3	108.7	97.2	105.9	96.9	74	100	0.053	0.087	-0.028	0.059	-0.031
85	98.5	86.9	98.3	104.3	113.3	100	100	-0.015	-0.131	-0.017	0.043	0.133
88	88.3	86.8	95.0	102.8	101.6	97	100	-0.117	-0.132	-0.050	0.028	0.016
90	88.3	109.2	99.9	103.7	80.9	95	100	-0.117	0.092	-0.001	0.037	-0.191
98	119.4	116.9	96.8	98.2	102.1	87	100	0.194	0.169	-0.032	0.018	0.021
107	102.8	107.7	95.6	100.3	95.4	110	100	0.028	0.077	-0.044	0.003	-0.046
111	111.1	102.6	98.8	101.9	108.3	122	100	0.111	0.026	-0.012	0.019	0.083
114	64.4	100.4	103.9	95.1	64.2	103	100	-0.356	0.004	0.039	0.049	-0.358
118	122.0	114.5	96.5	104.9	106.5	115	100	0.220	0.145	-0.035	0.049	0.065
124	135.2	128.3	96.8	101.5	105.4	109	100	0.352	0.283	-0.032	0.015	0.054

Chapter 5 Tables

(no tables)

BIBLIOGRAPHY

- Ablett, G.R., W.D. Beversdorf, and V.A. Dirks. 1991. Row spacing and seeding rate performance of indeterminate, semideterminate, and determinate soybean. *J. Prod. Agric.* 4:391–395.
- Ainsworth, E. A., Rogers, A., Nelson, R., and Long, S. P. 2004. Testing the “source-sink” hypothesis of down-regulation of photosynthesis in elevated [CO₂] in the field with single gene substitutions in *Glycine max.* *Agr. Forest Meteorol.* 122:85–94.
- Akao, S., and Kouchi, H. 1992. A supernodulating mutant isolated from soybean cultivar enrei. *Soil Sci. Plant Nutr.* 38:183–187.
- Anai, T. 2012. Potential of a mutant-based reverse genetic approach for functional genomics and molecular breeding in soybean. *Breeding Sci.* 61:462–467.
- Anai, T., Yamada, T., Hideshima, R., Kinoshita, T., Rahman, S. M., and Takagi, Y. 2008. Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by GmFAD2-1a. *Breeding Sci.* 58:447–452.
- Anders, S., and Huber, W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.
- Anders S., Pyl, P.T., and Huber, W. 2014. HTSeq--A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.
- Anderson, J. E., Kantar, M. B., Kono, T. Y., Fu, F., Stec, A. O., Song, Q., Cregan, P. B., Specht, J. E., Diers, B. W., Cannon, S. B., et al. 2014. A Roadmap for Functional Structural Variants in the Soybean Genome. *G3 (Bethesda)* 4:1307–1318.

- Andrade, F. H., Calvino, P., Cirilo, A., and Barbieri, P. 2002. Yield responses to narrow rows depend on increased radiation interception. *Agronomy Journal*, 94: 975-980.
- Arase, S., Hase, Y., Abe, J., Kasai, M., Yamada, T., Kitamura, K., Narumi, I., Tanaka, A., and Kanazawa, A. 2011. Optimization of ion-beam irradiation for mutagenesis in soybean: effects on plant growth and production of visibly altered mutants. *Plant Biotechnol.* 28:323–329.
- Atak, Ç., Alikamanoğlu, S., Açıık, L., and Canbolat, Y. 2004. Induced of plastid mutations in soybean plant (*Glycine max* L. Merrill) with gamma radiation and determination with RAPD. *Mutat. Res.* 556:35–44.
- Bart R.S., Chern M., Vega-Sánchez M.E., Canlas P., and Ronald P.C. 2010. Rice Snl6, a cinnamoyl-CoA reductase-like gene family member, is required for NH1-mediated immunity to *Xanthomonas oryzae* pv. *oryzae*. *PLoS Genet* 6:e1001123.
- Behrens, M. R., Mutlu, N., Chakraborty, S., Dumitru, R., Jiang, W. Z., Lavalley, B. J., Herman, P. L., Clemente, T. E., and Weeks, D. P. 2007. Dicamba resistance: enlarging and preserving biotechnology-based weed management strategies. *Science* 316:1185–1188.
- Beilstein, M., and Szymanski, D. 2004. Cytoskeletal requirements during Arabidopsis trichome development. *Plant Cytoskeleton. Cell Differ. Dev.* pp 265–289.
- Belfield, E. J., Gan, X., Mithani, A., Brown, C., Jiang, C., Franklin, K., Alvey, E., Wibowo, A., Jung, M., Bailey, K., et al. 2012. Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. *Genome Res.* 22:1306–1315.

- Bernard, R. L. 1975. The inheritance of appressed pubescence. *Soybean Genet. Newsl.* 2:34–36
- Bernard, R. L. 1975. Genetic stocks available. *Soyb. Genet. Newsl.* 2: 57–74.
- Bernard, R. L. 1976. Soybean Genetic Type Collection List. *Soybean. Genet. Newsl.* 3:62–67.
- Bernard, R. L., and Cremeens, C. R. 1988. Registration of “Williams 82” Soybean. *Crop Sci.* 28:1027–1028.
- Bernard, R.L. and Lindahl, D.A. 1972. Registration of Williams Soybean. *Crop Science*, 12: 716.
- Bernard, R. L., Nelson, R. L., and Cremeens, C. R. 1991. USDA soybean genetic collection: isoline collection. *Soybean Genet. Newsl.* 18:27–57.
- Bernard, R. L., and Singh, B. B. 1969. Inheritance of Pubescence Type in Soybeans: Glabrous, Curly, Dense, Sparse, and Puberulent. *Crop Sci.* 9:192–197.
- Bolon, Y. T., Haun, W. J., Xu, W. W., Grant, D., Stacey, M. G., Nelson, R. T., Gerhardt, D. J., Jeddeloh, J. A., Stacey, G., Muehlbauer, G. J., et al. 2011. Phenotypic and Genomic Analyses of a Fast Neutron Mutant Population Resource in Soybean. *Plant Physiol.* 156:240–253.
- Bolon, Y. T., Stec, A. O., Michno, J. M., Roessler, J., Bhaskar, P. B., Ries, L., Dobbels, A. A., Campbell, B. W., Young, N. P., Anderson, J. E., et al. 2014. Genome Resilience and Prevalence of Segmental Duplications Following Fast Neutron Irradiation of Soybean. *Genetics* 198:967–981.
- Brink, K., Chui, C., Cressman, R. F., Garcia, P., Henderson, N., Hong, B., Maxwell, C.

- A., Meyer, K., Mickelson, J., Stecca, K. L., et al. 2014. Molecular Characterization, Compositional Analysis, and Germination Evaluation of a High-Oleic Soybean Generated by the Suppression of FAD2-1 Expression. *Crop Sci.* 54:2160-2174.
- Brossman, G. D., and Wilcox, J. R. 1984. Induction of Genetic Variation for Oil Properties and Agronomic Characteristics of Soybean. *Crop Sci.* 24:783–787.
- Bubeck, D. M., Fehr, W. R., and Hammond, E. G. 1989. Inheritance of Palmitic and Stearic Acid Mutants of Soybean. *Crop Sci.* 29:652–656.
- Bullock, D., S. Khan, and A. Rayburn. 1998. Soybean yield response to narrow rows is largely due to enhanced early growth. *Crop Sci.* 38:1011–1016.
- Burnside, O. C., and Colville, W. L. 1964. Soybean and weed yields as affected by irrigation, row spacing, tillage, and amiben. *Weed Science*, 12: 109-112.
- De Bruin, J. L., and Pedersen, P. 2008. Effect of Row Spacing and Seeding Rate on Soybean Yield. *Agron. J.* 100: 704-710.
- Campbell, B. W., and Stupar, R. M. 2016. Soybean Mutant and Germplasm Resources: Current Status and Future Prospects. *Curr. Protoc. Plant Biol.* (in press).
- Campbell, B. W., Mani, D., Curtin, S. J., Slattery, R. A., Michno, J. M., Ort, D. R., Schaus, P. J., Palmer, R. G., Orf, J. H., and Stupar, R. M. 2014. Identical Substitutions in Magnesium Chelatase Paralogs Result in Chlorophyll Deficient Soybean Mutants. *G3 (Bethesda)* 5:123–131.
- Carpenter, J. A., and Fehr, W. R. 1986. Genetic Variability for Desirable Agronomic Traits in Populations Containing Glycine soja Germplasm. *Crop Sci.* 26:681–686.

- Carrero-Colón, M., Abshire, N., Sweeney, D., Gaskin, E., and Hudson, K. 2014. Mutations in SACPD-C result in a range of elevated stearic acid concentration in soybean seed. *PLoS One* 9:e97891.
- Carroll, B. J., McNeil, D. L., and Gresshoff, P. M. 1985a. A Supernodulation and Nitrate-Tolerant Symbiotic (nts) Soybean Mutant. *Plant Physiol.* 78:34–40.
- Carroll, B. J., McNeil, D. L., and Gresshoff, P. M. 1985b. Isolation and properties of soybean [*Glycine max* (L.) Merr.] mutants that nodulate in the presence of high nitrate concentrations. *Proc. Natl. Acad. Sci. U.S.A.* 82:4162–4166.
- Carroll, B. J., McNeil, D. L., and Gresshoff, P. M. 1986. Mutagenesis of soybean (*Glycine max* (L.) Merr.) and the isolation of non-nodulating mutants. *Plant Sci.* 47:109–114.
- Carter, T. E., Nelson, R., Sneller, C. H., and Cui, Z. 2004. Genetic diversity in soybean. In *Soybeans: Improvement, Production and Uses*, 3rd ed. H. R. Boerma and J. E. Specht, eds., pp. 303–416. ASA-CSSA-SSSA, Madison, WI.
- Cary, T.R., and Nickell, C.D. 1999. Genetic analysis of a short-petiolule-type soybean, LN89-3502TP. *J. Hered.* 90: 300-301.
- Chen, K., and Gao, C. 2014. Targeted genome modification technologies and their applications in crop improvements. *Plant Cell Rep.* 33:575–583.
- Cianzio, S. R., Shultz, S. P., Fehr, W. R., and Tachibana, H. 1991. Registration of ‘Archer’ soybean. *Crop Science* 31: 1707.
- Cleary, A. L., and Smith, L. G. 1998. The Tangled1 gene is required for spatial control of cytoskeletal arrays associated with cell division during maize leaf

- development. *Plant Cell* 10:1875-1888.
- Clough, S. J., and Bent, A. F. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735–743
- Cober, E. R., Molnar, S. J., Charette, M., and Voldeng, H. D. 2010. A new locus for early maturity in soybean. *Crop Sci.* 50:524–527.
- Concibido, V. C., La Vallee, B., McLaird, P., Pineda, N., Meyer, J., Hummel, L., Yang, J., Wu, K., and Delannay, X. 2003. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theor. Appl. Genet.* 106:575–582.
- Concibido, V. C., Lange, D. A., Denny, R. L., Orf, J. H., and Young, N. D. 1997. Genome Mapping of Soybean Cyst Nematode Resistance Genes in “Peking”, PI 90763, and PI 88788 Using DNA Markers. *Crop Sci.* 37:258–264.
- Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., Wang, J., Hughes, T. J., Willis, D. K., Clemente, T. E., et al. 2012. Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* 338:1206–1209.
- Cooper, R. L. 1977. Response of Soybean Cultivars to Narrow Rows and Planting Rates Under Weed-free Conditions. *Agron J* 69: 89-92.
- Cooper, R. L. 1981. Development of Short-Statured Soybean Cultivars. *Crop Sci.* 21: 127-131.
- Cooper, R. L., and Jeffers, D. L. 1984. Use of Nitrogen Stress to Demonstrate the Effect of Yield Limiting Factors on the Yield Response of Soybean to Narrow Row

- Systems. *Agron. J.* 76: 257-259.
- Cooper, J. L., Till, B. J., Laport, R. G., Darlow, M. C., Kleffner, J. M., Jamai, A., El-Mellouki, T., Liu, S., Ritchie, R., Nielsen, N., et al. 2008. TILLING to detect induced mutations in soybean. *BMC Plant Biol.* 8:9.
- Costa, J. A., Oplinger, E. S., and Pendleton, J. W. 1980. Response of Soybean Cultivars to Planting Patterns. *Agron. J.* 72: 153-156
- Courtial, B., Feuerbach, F., Eberhard, S., Rohmer, L., Chiapello, H., Camilleri, C., and Lucas, H. 2001. Tnt1 transposition events are induced by in vitro transformation of *Arabidopsis thaliana*, and transposed copies integrate into genes. *Mol. Genet. Genomics* 265:32–42.
- Cserzo, M., Eisenhaber, F., Eisenhaber, B., and Simon, I. 2002. On filtering false positive transmembrane protein predictions. *Protein Eng.* 15: 745-752.
- Cui, Y., Barampuram, S., Stacey, M. G., Hancock, C. N., Findley, S., Mathieu, M., Zhang, Z., Parrott, W. A., and Stacey, G. 2013. Tnt1 Retrotransposon Mutagenesis: A Tool for Soybean Functional Genomics. *Plant Physiol.* 161:36–47.
- Curtin, S. J., Michno, J. M., Campbell, B. W., Gil-Humanes, J., Mathioni, S. M., Hammond, R., Gutierrez-Gonzalez, J. J., Donohue, R. C., Kantar, M. B., Eamens, A. L., Meyers, B. C., Voytas, D. F., and Stupar, R. M. 2015. MicroRNA Maturation and MicroRNA Target Gene Expression Regulation Are Severely Disrupted in Soybean dicer-like1 Double Mutants. *G3 (Bethesda)* 6:423-433.

- Curtin, S. J., Zhang, F., Sander, J. D., Haun, W. J., Starker, C., Baltes, N. J., Reyon, D., Dahlborg, E. J., Goodwin, M. J., Coffman, A. P., et al. 2011. Targeted Mutagenesis of Duplicated Genes in Soybean with Zinc-Finger Nucleases. *Plant Physiol.* 156:466–473.
- d’Erfurth, I., Cosson, V., Eschstruth, A., Lucas, H., Kondorosi, A., and Ratet, P. 2003. Efficient transposition of the Tnt1 tobacco retrotransposon in the model legume *Medicago truncatula*. *Plant J.* 34:95–106.
- Darvasi, A., and Soller, M. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141:1199-1207.
- Deeks, M. J., Kaloriti, D., Davies, B., Malhó, R., and Hussey, P. J. 2004. Arabidopsis NAP1 is essential for Arp2/3-dependent trichome morphogenesis. *Curr Biol* 14:1410–1414.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498.
- Devlin, D. L., Fjell, D. L., Shroyer, J. P., Gordon, W. B., Marsh, B. H., Maddux, L. D., Martin, V. L., and Duncan, S. R. 1995. Row Spacing and Seeding Rates for Soybean in Low and High Yielding Environments. *J. Prod. Agric.* 8: 215-222
- Dierking, E. C., and Bilyeu, K. D. 2009. New sources of soybean seed meal and oil composition traits identified through TILLING. *BMC Plant Biol.* 9:89.
- Ehleringer, J. R., and Mooney, H. A. 1978. Leaf hairs: effects on physiological activity

- and adaptive value to a desert shrub. *Oecologia* 37:183–200.
- El-Assal, S. D., Le, J., Basu, D., Mallery, E.L., and Szymanski, D. B. 2004. Arabidopsis GNARLED Encodes a NAP125 Homolog that Positively Regulates ARP2/3. *Curr Biol* 14:1405–1409.
- Elmore, R.W. 1998. Soybean cultivar responses to row spacing and seeding rates in rainfed and irrigated environments. *J. Prod. Agric.* 11: 326-331.
- Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*, 2: 953-971.
- Evans, L.T. 1997. Adapting and improving crops: the endless task. *Philosophical Transactions of The Royal Society of London Series B-Biology Sciences* 352: 901–906.
- Falco, S. C., Guida, T., Locke, M., Mauvais, J., Sanders, C., Ward, R. T., and Webber, P. 1995. Transgenic Canola and Soybean Seeds with Increased Lysine. *Nat. Biotechnol.* 13:577–582.
- Fehr, W. R., Welke, G. A., Cianzio, S. R., Duvick, D. N., and Hammond, E. G. 1991. Inheritance of Reduced Palmitic Acid Content in Seed Oil of Soybean. *Crop Sci.* 31:88–89.
- Findley, S. D., Pappas, A. L., Cui, Y., Birchler, J. A., Palmer, R. G., and Stacey, G. 2011. Fluorescence in situ hybridization-based karyotyping of soybean translocation lines. *G3 (Bethesda)* 1:117-129.
- Fischerova, H. 1975. Linkage relationships of recessive chlorophyll mutations in *Arabidopsis thaliana*. *Biol. Plant* 17: 182–188.

- Fodje, M., Hansson, N., A., Hansson, M., Olsen, J. G., Gough, S. et al. 2001. Interplay between an AAA module and an integrin I domain may regulate the function of magnesium chelatase. *J. Mol. Biol.* 311: 111-122.
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531-1545.
- Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C., and Feldman, M. W. 2002. Evolutionary rate in the protein interaction network. *Science* 296: 750-752.
- Fusaro, A. F., Matthew, L., Smith, N. A., Curtin, S. J., Dedic-Hagan, J. et al. 2006. RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. *EMBO Rep.* 7: 1168-1175.
- Gibson, L. C., Willows, R. D., Kannangara, C. G., von Wettstein, D., and Hunter, C. N. 1995. Magnesium-protoporphyrin chelatase of *Rhodobacter sphaeroides*: reconstitution of activity by combining the products of the *bchH*, *-I*, and *-D* genes expressed in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 92: 1941-1944.
- Gidoni, D., Fuss, E., Burbidge, A., Speckmann, G., James, S., Nijkamp, D., Mett, A., Feiler, J., Smoker, M., de Vroomen, M. J., et al. 2003. Multi-functional T-DNA/Ds tomato lines designed for gene cloning and molecular and physical dissection of the tomato genome. *Plant Mol. Biol.* 51:83–98.
- Gifford, R.M., Thorne, J.H., Hitz, W.D. and Giaquinta, R.T. 1984. Crop productivity and photoassimilate partitioning. *Science*, 225: 801-808.
- Gillman, J. D., Baxter I., and Bilyeu, K. 2013. Phosphorus Partitioning of Soybean Lines

- Containing Different Mutant Alleles of Two Soybean Seed-Specific Adenosine Triphosphate-Binding Cassette Phytic Acid Transporter Paralogs. *Plant Genome* doi:10.3835/plantgenome2012.06.0010.
- Gillman, J. D., Stacey, M. G., Cui, Y., Berg, H. R., and Stacey, G. 2014. Deletions of the SACPD-C locus elevate seed stearic acid levels but also result in fatty acid and morphological alterations in nitrogen fixing nodules. *BMC Plant Biol.* 14:143.
- Gizlice, Z., Carter, T. E., and Burton, J. W. 1994. Genetic Base for North American Public Soybean Cultivars Released between 1947 and 1988. *Crop Sci.* 34:1143–1151.
- Gobinath, P., and Pavadai, P. 2015. Morphology and Yield parameters and Biochemical analysis of Soybean (*Glycine max* (L.) Mrr.) Using Gamma rays, EMS and DES treatment. *Int. Lett. Nat. Sci.* 35:50–58.
- Gong, J., Waner, D. A., Horie, T., Li, S. L., Horie, R., Abid, K. B., and Schroeder, J. I. 2004. Microarray-based rapid cloning of an ion accumulation deletion mutant in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 101:15404–15409.
- Grandbastien, M. A., Spielmann, A., and Caboche, M. 1989. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* 337:376–380.
- Grant, D., Nelson, R. T., Cannon, S. B., and Shoemaker, R. C. 2010. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucl. Acids Res.* 38:D843-D846.
- Greer, H. A. L., and Anderson, I. C. 1965. Response to triiodobenzoic acid under field

- conditions. *Crop Sci.* 5:229-232.
- Gremaud, M. F., and Harper, J. E. 1989. Selection and initial characterization of partially nitrate tolerant nodulation mutants of soybean. *Plant Physiol.* 89:169–173.
- Groose, R. W., Schulte, S. M., and Palmer, R. G. 1990. Germinal reversion of an unstable mutation for anthocyanin pigmentation in soybean. *Theor. Appl. Genet.* 79:161–167.
- Groose, R. W., Weigelt, H. D., and Palmer, R. G. 1988. Somatic Analysis of an Unstable Mutation for Anthocyanin Pigmentation in Soybean. *J. Hered.* 79:263–267.
- Guo, J., Wang, Y., Song, C., Zhou, J., Qiu, L., Huang, H., and Wang, Y. 2010. A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. *Ann. Bot.* 106:505–514.
- Haase, N. J., Beissinger, T., Hirsch, C. N., Vaillancourt, B., Deshpande, S., Barry, K., Buell, C. R., Kaeppler, S. M., and de Leon, N. 2015. Shared Genomic Regions Between Derivatives of a Large Segregating Population of Maize Identified Using Bulk Segregant Analysis Sequencing and Traditional Linkage Analysis. *G3 (Bethesda)* 5:1593–1602.
- Hajika, M., Igita, K., and Kitamura, K. 1991. A Line lacking all the Seed Lipoxygenase Isozymes in Soybean (*Glycine max* (L.) MERRILL) Induced by gamma-ray Irradiation. *Jpn. J. Breed.* 41:507–509.
- Hatfield, J. L., and Carlson, R. E. 1978. Photosynthetically active radiation, CO₂ uptake, and stomatal diffusive resistance profiles within soybean canopies. *Agron. J.* 70: 592-596.
- Hammond, E. G., and Fehr, W. R. 1975. Oil Quality Improvement in Soybeans-*Glycine*

- max* (L.) Merr. *Fett. Wiss. Technol.* 77:97–101.
- Hammond, E. G., and Fehr, W. R. 1983. Registration of A5 Germplasm Line of Soybean (reg. No. GP44). *Crop Sci.* 23:192–193.
- Hammond, E. G., and Fehr, W. R. 1984. Improving the fatty acid composition of soybean oil. *J. Am. Oil Chem. Soc.* 61:1713–1716.
- Hancock, C. N., Zhang, F., Floyd, K., Richardson, A. O., LaFayette, P., Tucker, D., Wessler, S. R., and Parrott, W. A. 2011. The Rice Miniature Inverted Repeat Transposable Element mPing is an Effective Insertional Mutagen in Soybean. *Plant Physiol* 157:552–562.
- Hancock, C. N., Zhang, F., and Wessler, S. R. 2010. Transposition of the Tourist-MITE mPing in yeast: an assay that retains key features of catalysis by the class 2 PIF/Harbinger superfamily. *Mob. DNA* 1:5.
- Hansson, A., Kannangara, C. G., von Wettstein, D., and Hansson, M. 1999. Molecular basis for semidominance of missense mutations in the XANTHA-H (42-kDa) subunit of magnesium chelatase. *Proc. Natl. Acad. Sci. U.S.A.* 96: 1744-1749.
- Hansson, A., Willows, R. D., Roberts, T. H., and Hansson, M. 2002. Three semidominant barley mutants with single amino acid substitutions in the smallest magnesium chelatase subunit form defective AAA+ hexamers. *Proc. Natl. Acad. Sci. U.S.A.* 99: 13944-13949.
- Hartwig, E. E., and Epps, J. M. 1973. Registration of Forrest Soybeans. *Crop Science* 13:287.
- Haun, W. J., Hyten, D. L., Xu, W. W., Gerhardt, D. J., Albert, T. J., Richmond, T.,

- Jeddeloh, J. A., Jia, G., Springer, N. M., Vance, C. P., et al. 2011. The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. *Plant Physiol.* 155:645–655.
- Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., Retterath, A., Stoddard, T., Juillerat, A., Cedrone, F., et al. 2014. Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol. J.* 12:934–940.
- Hay, R. K. M. 1995. Harvest index: a review of its use in plant breeding and crop physiology. *Annals of applied biology* 126: 197-216.
- Healy, R. A., Horner, H. T., Bailey, T. B., and Palmer, R. G. 2005. A Microscopic Study of Trichomes on Gynoecia of Normal and Tetraploid Clark Cultivars of Glycine Max and Seven Near-Isogenic Lines. *Int. J. Plant Sci.* 166:415–425.
- Hicks, D. R., Pendleton, J. W., Bernard, R. L. and Johnston, T. J. 1969. Response of soybean plant types to planting patterns. *Agron. J.* 61: 290-293.
- Hirochika, H. 1993. Activation of tobacco. *EMBO J.* 12:2521–2528.
- Hitz, W. D., Carlson, T. J., Kerr, P. S., and Sebastian, S. A. 2002. Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiol.* 128:650–660.
- Horak, M. J., Rosenbaum, E. W., Kendrick, D. L., Sammons, B., Phillips, S. L., Nickson, T. E., Dobert, R. C., and Perez, T. 2015. Plant characterization of Roundup Ready 2 Yield ® soybean, MON 89788, for use in ecological risk assessment.

Transgenic Res. 24:213–225.

Hoshino, T., Watanabe, S., Takagi, Y., and Anai, T. 2014. A novel GmFAD3-2a mutant allele developed through TILLING reduces α -linolenic acid content in soybean seed oil. *Breed Sci.* 64:371-377.

Huang, Y. S., and Li, H. M. 2009. Arabidopsis CHLI2 can substitute for CHLI1. *Plant Physiol.* 150: 636-645.

Hulburt, D. J., Boerma, H. R., All, J. N. 2004. Effect of pubescence tip on soybean resistance to lepidopteran insects. *J. Econ. Entomol.* 97:621–627.

Humphrey, L. M. 1951. Effects of neutron irradiation on soybeans. *Soybean Digest.* 12:11-12.

Hymowitz, T., and Shurtleff, W. R. 2005. Debunking soybean myths and legends in the historical and popular literature. *Crop Sci.* 45:473–476.

Hymowitz, T., Orf, J. H., Kaizuma, N., and Skorupska, H. 1978. Screening the USDA soybean germplasm collection for Kunitz trypsin inhibitor variants. *Soybean Genet. Newsl.* 5:19–21.

Hyten, D. L., Q. Song, I. Y. Choi, M. S. Yoon, J. E. Specht et al. 2008. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theor. Appl. Genet.* 116: 945-952.

Hyten, D. L., Smith, J. R., Frederick, R. D., Tucker, M. L., Song, Q., and Cregan, P. B. 2009. Bulk Segregant Analysis Using the GoldenGate Assay to Locate the Locus that Confers Resistance to Soybean Rust in Soybean. *Crop Sci.* 49:265–271.

- Hyten, D. L., Song, Q., Zhu, Y., Choi, I. Y., Nelson, R. L., Costa, J. M., Specht, J. E., Shoemaker, R. C., and Cregan, P. B. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl. Acad. Sci. U.S.A.* 103:16666–16671.
- Iantcheva, A., Chabaud, M., Cosson, V., Barascud, M., Schutz, B., Primard-Brisset, C., Durand, P., Barker, D. G., Vlahova, M., and Ratet, P. 2009. Osmotic shock improves Tnt1 transposition frequency in *Medicago truncatula* cv Jemalong during in vitro regeneration. *Plant Cell Rep.* 28:1563–1572.
- Illa-Berenguer, E., Van Houten, J., Huang, Z., and van der Knaap, E. 2015. Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. *Theor. Appl. Genet.* 128:1329–1342.
- Ininda, J., Fehr, W. R., Cianzio, S. R., and Schnebly, S. R. 1996. Genetic Gain in Soybean Populations with Different Percentages of Plant Introduction Parentage. *Crop Sci.* 36:1470–1472.
- Izawa, T., Ohnishi, T., Nakano, T., Ishida, N., Enoki, H., Hashimoto, H., Itoh, K., Terada, R., Wu, C., Miyazaki, C., et al. 1997. Transposon tagging in rice. *Plant Mol. Biol.* 35:219–229.
- Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., and Parrott, W. A. 2015. Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* 15:16.
- James, G., Patel, V., Nordström, K. J., Klasen, J. R., Salomé, P. A., Weigel, D., and Schneeberger, K. 2013. User guide for mapping-by-sequencing in Arabidopsis. *Genome Biol.* 14:R61.
- Jensen, P. E., Willows, R. D., Petersen, B. L., Vothknecht, U. C., Stummann, B. M. et al.

1996. Structural genes for Mg-chelatase subunits in barley: Xantha-f, -g and -h. *Mol. Gen. Genet.* 250: 383-394.
- Jeong, N., Suh, S.J., Kim, M.H., Lee, S., Moon, J.K., Kim, H.S. and Jeong, S.C. 2012. Ltn is a key regulator of leaflet shape and number of seeds per pod in soybean. *Plant Cell* 24: 4807-4818.
- Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S. R., McCouch, S. R., and Wessler, S. R. 2003. An active DNA transposon family in rice. *Nature* 421:163–167.
- Jiang, G. M., Sun, J. Z., Liu, H. Q., Qu, C. M., Wang, K. J., Guo, R. J., Bai, K. Z., Gao, L. M., and Kuang, T. Y. 2003. Changes in the rate of photosynthesis accompanying the yield increase in wheat cultivars released in the past 50 years. *Journal of plant research*, 116: 347-354.
- Johnson, C. S. 2002. TRANSPARENT TESTA GLABRA2, a Trichome and Seed Coat Development Gene of Arabidopsis, Encodes a WRKY Transcription Factor. *Plant Cell Online* 14:1359–1375.
- Johnson, H. W. 1958. Registration of soybean varieties. *Agron. J.* 50:690–691
- Johnson, H. W., and Hollowell, E. A. 1935. Pubescent and glabrous characters of soybeans as related to resistance to injury by the potato leafhopper. *J. Agric. Res.* 51:371–381.
- Johnston, T. J., Pendleton, J. W., Peters, D. B., and Hicks, D. R. 1969. Influence of supplemental light on apparent photosynthesis, yield, and yield components of soybeans (*Glycine max* L.). *Crop Sci.* 9: 577-581.
- Jones, J. D., Carland, F., Lim, E., Ralston, E., and Dooner, H. K. 1990. Preferential

- transposition of the maize element Activator to linked chromosomal locations in tobacco. *Plant Cell* 2:701–707.
- Jun, T.H., Kang, S.T., Moon, J.K., Seo, M.J., Yun, H.T., Lee, S.K., Lee, Y.H. and Kim, S.J. 2009. Genetic analysis of new short petiole gene in soybean. *J. Crop Sci. Biotechnol.* 12: 87-89.
- Jung, K. H., J. Hur, C. H. Ryu, Y. Choi, Y. Y. Chung et al. 2003. Characterization of a rice chlorophyll-deficient mutant using the T-DNA gene-trap system. *Plant Cell Physiol.* 44: 463-472.
- Kabelka, E. A., Carlson, S. R., and Diers, B. W. 2006. PI 468916 SCN Resistance Loci's Associated Effects on Soybean Seed Yield and Other Agronomic Traits. *Crop Sci.* 46:622–629.
- Kato, K. K., and Palmer, R.G. 2004. Duplicate chlorophyll-deficient loci in soybean. *Genome* 47: 190-198.
- Kelley, L. A., and Sternberg, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* 4: 363-371.
- Kikuchi, K., Terauchi, K., Wada, M., and Hirano, H. Y. 2003. The plant MITE mPing is mobilized in anther culture. *Nature* 421:167–170.
- Kilen, T.C. 1983. Inheritance of a short petiole trait in soybean. *Crop Sci.* 23: 1208–1209
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36.
- Kim, D. S., Lee, K. J., Kim, J. B., Kim, S. H., Song, J. Y., Seo, Y. W., Lee, B. M., and

- Kang, S. Y. 2010. Identification of Kunitz trypsin inhibitor mutations using SNAP markers in soybean mutant lines. *Theor. Appl. Genet.* 121:751–760.
- Kim, M. Y., Lee, S., Van, K., Kim, T. H., Jeong, S. C., Choi, I. Y., et al. 2010. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc. Natl. Acad. Sci. U.S.A.* 107:22032-22037.
- Kinney, A. J. 1996. Development of genetically engineered soybean oils for food applications. *J. Food Lipids* 3:273–292.
- Kitagawa, S., Ishimoto, M., Kikuchi, F., and Kitamura, K. 1991. A characteristic lacking or decreasing remarkably 7S globulin subunits induced with γ -ray irradiation in soybean seeds. *Jpn. J. Breed.* 41:460–461.
- Kobayashi, K., N. Mochizuki, N. Yoshimura, K. Motohashi, T. Hisabori et al. 2008. Functional analysis of *Arabidopsis thaliana* isoforms of the Mg-chelatase CHLI subunit. *Photochem. Photobiol. Sci.* 7: 1188-1195.
- Kokubun, M. 1988. Design and evaluation of soybean ideotypes. *Bulletin of the Tohoku National Ag. Exp. Station.*
- Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C. S., Jiang, S., Ramamoorthy, R., Cai, M., Ma, Z. G., Sundaresan, V., and Ramachandran, S. 2004. Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. *Plant J.* 37:301–314.
- Konieczny, A., and Ausubel, F. M. 1993. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *The Plant Journal* 4: 403-410.

- Koncz, C., Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Reiss, B., et al. 1990. Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana*. *EMBO J.* 9: 1337-1346.
- Kumari, S., Lal, S. K., and Sachdev, A. 2014. Identification of putative low phytic acid mutants and assessment of the total P, phytate P, protein and divalent cations in mutant populations of soybean. *Aust. J. Crop Sci.* 8:435–441.
- Lee, H. S., Chae, Y. A., Park, E. H., Kim, Y. W., Yun, K. I., and Lee, S. H. 1997. Introduction, development, and characterization of supernodulating soybean mutant. I. Mutagenesis of soybean and selection of supernodulating mutant. *Korean J. Crop Sci.* 42:247–253.
- Lee, J. D., Shannon, J. G., Vuong, T. D., and Nguyen, H. T. 2009. Inheritance of salt tolerance in wild soybean (*Glycine soja* Sieb. and Zucc.) accession PI483463. *J. Hered.* 100:798–801.
- Levin, D. A. 1973. The Role of Trichomes in Plant Defense. *Q Rev Biol* 48:3–15
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li, Y. H., Liu, Y. L., Reif, J. C., Liu, Z. X., Liu, B., Mette, M. F., Chang, R. Z., and Qiu, L. J. 2014. Biparental Resequencing Coupled With SNP Genotyping of a Segregating Population Offers Insights Into the Landscape of Recombination and

- Fixed Genomic Regions in Elite Soybean. *G3 (Bethesda)* 4:553–560.
- Li, D., Pfeiffer, T. W., and Cornelius, P. L. 2008. Soybean QTL for yield and yield components associated with Glycine soja alleles. *Crop Sci.* 48:571–581.
- Li, X., Song, Y., Century, K., Straight, S., Ronald, P., Dong, X., Lassner, M., and Zhang, Y. 2001. A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant J.* 27:235–242.
- Li, Y., Zhao, S., Ma, J., Li, D., Yan, L., Li, J., Qi, X., Guo, X., Zhang, L., He, W., et al. 2013. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics* 14:579.
- Li, Z., Liu, Z. B., Xing, A., Moon, B. P., Koellhoffer, J. P., Huang, L., Ward, R. T., Clifton, E., Falco, S. C., and Cigan, A. M. 2015. Cas9-Guide RNA Directed Genome Editing in Soybean. *Plant Physiol.* 169:960–970.
- Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G., and Stacey, G. 2010. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J* 63: 86-99.
- Lichtenthaler, H. K. 1987. Chlorophylls and carotenoids - pigments of photosynthetic biomembranes. *Methods Enzymol.* 148: 350-382.
- Lightner, J., and Caspar, T. 1998. Seed Mutagenesis of Arabidopsis. In *Arabidopsis Protocols* pp. 91–102. Humana Press, Totowa, NJ.
- Liu, S., Yeh, C. T., Tang, H. M., Nettleton, D., and Schnable, P. S. 2012. Gene Mapping via Bulk Segregant RNA-Seq (BSR-Seq). *PLoS One* 7:e36406.

- Liu Y. G., Mitsukawa, N., Oosumi, T., and Whittier, R. F. 1995. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* 8:457–463.
- Liu, B., Fujita, T., Yan, Z. H., Sakamoto, S., Xu, D., and Abe, J. 2007. QTL mapping of domestication-related traits in soybean (*Glycine max*). *Ann. Bot.* 100:1027–1038.
- Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., Nagamatsu, A., Arai, M., Yamada, T., Kitamura, K. and Masuta, C. 2010. The soybean stem growth habit gene Dt1 is an ortholog of *Arabidopsis* TERMINAL FLOWER1. *Plant Physiol.* 153:198-210.
- Liu, S., Kandath, P. K., Warren, S. D., Yeckel, G., Heinz, R., Alden, J., Yang, C., Jamai, A., El-Mellouki, T., Juvale, P. S., et al. 2012. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* 492:256-260.
- Liu, S., Kandath, P. K., Warren, S. D., Yeckel, G., Heinz, R., et al. 2012. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* 492: 256-260.
- Long, S. P., Zhu, X. G., Naidu, S. L., and Ort, D. R. 2006. Can improvement in photosynthesis increase crop yields?. *Plant, Cell & Environment* 29:315-330.
- Loomis, R. S., Williams, W. A., Duncan, W. G., San Pietro, A., Greer, F., and Army, T. 1967. Community architecture and the productivity of terrestrial plant communities. *Harvesting the sun. Photosynthesis in plant life.*, 291-308.
- Lorenz, A.J., Gustafson, T.J., Coors, J.G. and Leon, N.D. 2010. Breeding maize for a

- bioeconomy: a literature survey examining harvest index and stover yield and their relationship to grain yield. *Crop Science*, 50:1-12.
- Lucas, H., Feuerbach, F., Kunert, K., Grandbastien, M. A., and Caboche, M. 1995. RNA-mediated transposition of the tobacco retrotransposon Tnt1 in *Arabidopsis thaliana*. *EMBO J.* 14:2364–2373.
- Lueschen, W.E., Ford, J.H., Evans, S.D., Kanne, B.K., Hoverstad, T.R., Randall, G.W., Orf, J.H., and Hicks, D.R. 1992. Tillage, row spacing, and planting date effects on soybean following corn and wheat. *J. Prod. Agric.* 5:254–260.
- Lynch, M., O'Hely, M., Walsh, B., and Force, A. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* 159: 1789-1804.
- Macdonald, S. J. and Long, A. D. 2007. Joint estimates of quantitative trait locus effect and frequency using synthetic recombinant populations of *Drosophila melanogaster*. *Genetics* 176: 1261–1281.
- Mahama, A. A., and Palmer, R. G. 2003. Translocation breakpoints in soybean classical genetic linkage groups 6 and 8. *Crop Sci.* 43: 1602-1609.
- Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., et al. 2013. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* 41: D348-D352.
- Marks, M. D., Wenger, J. P., Gilding, E., Jilk, R., and Dixon, R. A. 2009. Transcriptome analysis of *Arabidopsis* wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. *Mol. Plant* 2:803-822.
- Martienssen, R. A. 1998. Functional genomics: probing plant gene function and

- expression with transposons. *Proc. Natl. Acad. Sci. U.S.A.* 95:2021–2026.
- Mascher, M., Jost, M., Kuon, J.E., Himmelbach, A., Aßfalg, A., Beier, S., Scholz, U., Graner, A., and Stein, N. 2014. Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol.* 15:R78.
- Mathieu, M., Winters, E. K., Kong, F., Wan, J., Wang, S., Eckert, H., Luth, D., Paz, M., Donovan, C., Zhang, Z., et al. 2009. Establishment of a soybean (*Glycine max* Merr. L) transposon-based mutagenesis repository. *Planta* 229:279–289.
- Mathur, J., Mathur, N., Kirik, V., Kernebeck, B., Srinivas, B. P., and Hülskamp, M. 2003. Arabidopsis CROOKED encodes for the smallest subunit of the ARP2/3 complex and controls cell shape by region specific fine F-actin formation. *Development* 130:3137–3146.
- Mazier, M., Botton, E., Flamain, F., Bouchet, J. P., Courtial, B., Chupeau, M. C., Chupeau, Y., Maisonneuve, B., and Lucas, H. 2007. Successful Gene Tagging in Lettuce Using the Tnt1 Retrotransposon from Tobacco. *Plant Physiol.* 144:18–31.
- McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. 2000. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123:439–442.
- McHale, L. K., Haun, W. J., Xu, W. W., Bhaskar, P. B., Anderson, J. E., Hyten, D. L., Gerhardt, D. J., Jeddeloh, J. A., and Stupar, R. M. 2012. Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol* 159:1295–1308.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A.,

- Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.
- Men, A. E., Laniya, T. S., Searle, I. R., Iturbe-Ormaetxe, I., Gresshoff, I., Jiang, Q., Carroll, B. J., and Gresshoff, P. M. 2002. Fast Neutron Mutagenesis of Soybean (*Glycine soja* L.) Produces a Supernodulating Mutant Containing a Large Deletion in Linkage Group H. *Genome Letters* 1:147–155.
- Michelmore, R. W., Paran, I., and Kesseli, R. V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. U.S.A.* 88: 9828-9832.
- Michno, J. M., Wangab, X., Liu, J., Curtin, S. J., Kono, T. J. Y., and Stupar, R. M. 2015. CRISPR/Cas mutagenesis of soybean and *Medicago truncatula* using a new web-tool and a modified Cas9 enzyme. *GM Crops Food* (in press.) doi: 10.1080/21645698.2015.1106063.
- Mochizuki, N., Brusslan, J. A., Larkin, R., Nagatani, A., and Chory, J. 2001. Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proc. Natl. Acad. Sci. U.S.A.* 98: 2053-2058.
- Morrell, P. L., Toleno, D. M., Lundy, K. E., and Clegg, M. T. 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proc. Natl. Acad. Sci. U.S.A.* 102:2442–2447.

- Morrison, M. J., Voldeng, H.D. and Cober, R.R. 1999. Physiological changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. *Agron. J.* 91: 685–689.
- Morrison, M. J., Voldeng, H.D. and Cober, R.R. 2000. Agronomic changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. *Agron. J.* 92: 780–784.
- Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C., and Flint, J. 2000. A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl. Acad. Sci. U.S.A* 97: 12649-12654.
- Mounts, Y. L., Warner, K., Lista, G. R., Kleiman, R., Fehro, W. R., Hammondo, E. G., and Wileoxc, J. R. 1988. Effect of Altered Fatty Acid Composition on Soybean Oil Stability. *J. Am. Oil Chem. Soc.* 65:624–628.
- Mrocza, A., Roberts, P. D., Fillatti, J. J., Wiggins, B. E., Ulmasov, T., and Voelker, T. 2010. An Intron Sense Suppression Construct Targeting Soybean FAD2-1 Requires a Double-Stranded RNA-Producing Inverted Repeat T-DNA Insert. *Plant Physiol.* 153:882–891.
- Nagai, I., and Saito, S. 1923. Linked factors in soybeans. *Japan Jour Bot* 1:121–136
- Naito, K., Cho, E., Yang, G., Campbell, M. A., Yano, K., Okumoto, Y., Tanisaka, T., and Wessler, S. R. 2006. Dramatic amplification of a rice transposable element during recent domestication. *Proc. Natl. Acad. Sci. U.S.A.* 103:17620–17625.
- Nakayama, M., Masuda, T., Sato, N., Yamagata, H., Bowler, C., et al. 1995. Cloning, subcellular localization and expression of CHL1, a subunit of magnesium-

- chelataase in soybean. *Biochem. Biophys. Res. Commun.* 215: 422-428.
- Nakazaki, T., Okumoto, Y., Horibata, A., Yamahira, S., Teraishi, M., Nishida, H., Inoue, H., and Tanisaka, T. 2003. Mobilization of a transposon in the rice genome. *Nature* 421:170–172.
- Nei, M., and Li, W. H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* 76:5269–5273.
- Nguyen, L. V., 1995. Transposon tagging and isolation of the sulfur gene in tobacco (*Nicotiana tabacum*). Ph.D. thesis, North Carolina State University, Raleigh, North Carolina.
- Nichols, D. M., Glover, K. D., Carlson, S. R., Specht, J. E., and Diers, B. W. 2006. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. *Crop Sci.* 46:834–839.
- Nyman, Y. 1993. The pollen-collecting hairs of *Campanula* (Campanulaceae). I. Morphological variation and the retractive mechanism. *Am J Bot* 80:1427–1436.
- O'Rourke, J. A., Iniguez, L. P., Fu, F., Bucciarelli, B., Miller, S. S., Jackson, S. A., McClean, P. E., Li, J., Dai, X., Zhao, P. X., et al. 2014. An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics* 15:866.
- Odanaka, H., and Kaizuma, N. 1989. Mutants on soybean storage proteins induced with γ -ray irradiation. *Jpn. J. Breed.* 39:430–431.
- Oplinger, E.S., and B.D. Philbrook. 1992. Soybean planting date, row width, and seeding rate response in three tillage systems. *J. Prod. Agric.* 5:94–99.
- Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., and Marks, M. D. 1991.

- A myb gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. *Cell* 67:483–493.
- Oppenheimer, D. G., Pollock, M. A., Vacik, J., Szymanski, D. B., Ericson, B., Feldmann, K., and Marks, M. D. 1997. Essential role of a kinesin-like protein in Arabidopsis trichome morphogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 94:6261–6266.
- Orf, J. H., and Denny, R. L. 2004. Registration of “MN1302” Soybean. *Crop Sci.* 44:693.
- Orf, J. H., and Hymowitz, T. 1979. Inheritance of the Absence of the Kunitz Trypsin Inhibitor in Seed Protein of Soybeans. *Crop Sci.* 19:107–109.
- Orf, J. H., Lambert, J. W., and Kennedy, B. W. 1985. Registration of Ozzie soybean. *Crop Science* 25: 366-366.
- Owen, F. V. 1927. Inheritance studies in soybeans. II. Glabrousness, color of pubescence, time of maturity, and linkage relations. *Genetics* 12:519–529.
- Padgett, S. R., Kolacz, K. H., Delannay, X., Re, D. B., LaVallee, B. J., Tinius, C. N., Rhodes, W. K., Otero, Y. I., Barry, G. F., Eichholtz, D. A., et al. 1995. Development, identification, and characterization of a glyphosate-tolerant soybean line. *Crop Sci.* 35:1451–1461.
- Palmer, R. G., Hedges, B. R., Benavente, R. S., and Groose, R. W. 1989. w4-Mutable line in soybean. *Dev. Genet.* 10:542–551.
- Palmer, R. G., R. L. Nelson, R. L. Bernard, and Stelly D. M. 1990. Genetics and Linkage of 3 Chlorophyll-Deficient Mutants in Soybean - Y19, Y22, and Y23. *J. Hered.* 81: 404-406.
- Palmer, R. G., Sandhu, D., Curran, K., and Bhattacharyya, M. K. 2008a. Molecular

- mapping of 36 soybean male-sterile, female-sterile mutants. *Theor. Appl. Genet.* 117:711–719.
- Palmer, R. G., T. W. Pfeiffer, G. R. Buss, and Kilen, T. C. 2004. Qualitative genetics, pp. 137-233 in *Soybeans: Improvement, Production, and Uses*, Third Ed., edited by H. R. Boerma and J. E. Specht, ASA-CSSA-SSSA, Madison WI.
- Palmer, R. G., Zhang, L., Huang, Z. P., and Xu, M. 2008b. Allelism and molecular mapping of soybean necrotic root mutants. *Genome* 51:243–250.
- Pendleton, J. W., Smith, G. E., Winter, S. R., and Johnston, T. J. 1968. Field investigations of the relationships of leaf angle in corn (*Zea mays* L.) to grain yield and apparent photosynthesis. *Agron. J.* 60: 422-424.
- Pham, A. T., Lee, J. D., Shannon, J. G., and Bilyeu, K. D. 2010. Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. *BMC Plant Biol.* 10:195.
- Ping, J., Liu, Y., Sun, L., Zhao, M., Li, Y., She, M., Sui, Y., Lin, F., Liu, X., Tang, Z. and Nguyen, H. 2014. Dt2 is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean. *Plant Cell* 26: 2831-2842.
- Piper, C. V., and Morse, W.J. 1923. *The Soybean*. McGraw-Hill Book Co, New York
- Porra, R. J., W. A. Thompson, and Kriedemann, P. E. 1989. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: Verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta* 975: 384-394.

- R Development Core Team. 2006. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, ISBN 3-900051-07-0.
- Rahman, S. M., Kinoshita, T., Anai, T., Arima, S., and Takagi, Y. 1998. Genetic Relationships of Soybean Mutants for Different Linolenic Acid Contents. *Crop Sci.* 38:702–706.
- Rahman, S. M., Takagi, Y., Kubota, K., Miyamoto, K., and Kawakita, T. 1994. High Oleic Acid Mutant in Soybean Induced by X-Ray Irradiation. *Biosci. Biotech. Bioch.* 58:1070–1072.
- Rahman, S. M., Takagi, Y., Miyamoto, K., and Kawakita, T. 1995. High Stearic Acid Soybean Mutant Induced by X-ray Irradiation. *Biosci. Biotech. Bioch.* 59:922–923.
- Raval, J., Baumbach, J., Ollhoff, A. R., Pudake, R. N., Palmer, R. G., Bhattacharyya, M. K., and Sandhu, D. 2013. A candidate male-fertility female-fertility gene tagged by the soybean endogenous transposon, Tgm9. *Funct. Integr. Genomics* 13:67–73.
- Rawlings, J. O., Hanway, D. G., and Gardner, C. O. 1958. Variation in Quantitative Characters of Soybeans After Seed Irradiation. *Agron. J.* 50:524-528.
- Rédei, G. P., Acedo, G. N., and Sandhu, S. S. 1984. Mutation Induction and Detection in Arabidopsis. In *Mutation, Cancer, and Malformation* pp. 285–313. Springer US, Boston, MA.
- Rerie, W.G., Feldmann, K.A., and Marks, M.D. 1994. The GLABRA2 gene encodes a

- homeo domain protein required for normal trichome development in Arabidopsis. *Genes Dev.* 8:1388–1399
- Riggs, T. J., Hanson, P. R., Start, N. D., Miles, D.M., Morgan, C.L., and Ford, M.A. 1981. Comparison of spring barley varieties grown in England and Wales between 1880 and 1980. *Journal of Agricultural Science, Cambridge* 97: 599–610.
- Riggs, R., Wang, S., Singh, R., and Hymowitz, T. 1998. Possible transfer of resistance to *Heterodera glycines* from *Glycine tomentella* to soybean. *J. Nematol.* 30:547–552.
- Ríos, G., Naranjo, M. A., Iglesias, D. J., Ruiz-Rivero, O., Geraud, M., Usach, A., and Talón, M. 2008. Characterization of hemizygous deletions in citrus using array-comparative genomic hybridization and microsynteny comparisons with the poplar genome. *BMC Genomics* 9:381.
- Rissler, H. M., Collakova, E., DellaPenna, D., Whelan, J., and Pogson, B. J. 2002. Chlorophyll biosynthesis. Expression of a second chl I gene of magnesium chelatase in Arabidopsis supports only limited chlorophyll synthesis. *Plant Physiol.* 128: 770-779.
- Robbins, J.C., Daugherty, D.M., and Hatchett, J.H. 1979. Ovipositional and Feeding Preference of Leafhoppers (Homoptera : Cicadellidae) on Clark Soybeans in Relation to Plant Pubescence. *J Kansas Entomol Soc* 52:603–608.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*

26:139–140.

- Rockman, M. V. and Kruglyak, L. 2008. Breeding designs for recombinant inbred advanced intercross lines. *Genetics* 179: 1069–1078.
- Ross, J. P., and Brim, C. A. 1957. Resistance of soybeans to the soybean cyst nematode as determined by a double-row method. *Plant Dis. Rep.* 41:923–924.
- Rotundo, J. L., Borrás, L., De Bruin, J., and Pedersen, P. 2012. Physiological strategies for seed number determination in soybean: Biomass accumulation, partitioning and seed set efficiency. *Field Crops Research* 135: 58-66.
- Ruddle, P., Whetten, R., Cardinal, A., Upchurch, R. G., and Miranda, L. 2013. Effect of a novel mutation in a $\Delta 9$ -stearoyl-ACP-desaturase on soybean seed oil composition. *Theor. Appl. Genet.* 126:241–249.
- Ryan, S. A., and Harper, J. E. 1983. Mutagenesis of soybeans. *Soybean Genet. Newsl.* 10:29–32.
- Sakamoto, C. M., and Shaw, R. H. 1967a. Light distribution in field soybean canopies. *Agron. J.* 59: 7-9.
- Sakamoto, C. M., and Shaw, R. H. 1967b. Apparent photosynthesis in field soybean communities. *Agron. J.* 59: 73-75.
- Sarlikioti, V. de Visser, P. H. B., Buck-Sorlin, G. H., and Marcelis, L. F. M. 2011. How plant architecture affects light absorption and photosynthesis in tomato: towards an ideotype for plant architecture using a functional-structural plant model. *Ann Bot* 108:1065–1073.
- Sawers, R. J., Viney, J., Farmer, P. R., Bussey, R. R., Olsefski, G. et al., 2006. The maize Oil yellow1 (Oy1) gene encodes the I subunit of magnesium chelatase.

Plant Mol. Biol. 60: 95-106.

- Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., Thumfahrt, J., Jürgens, G., and Hülskamp, M. 2002. TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. *EMBO J* 21:5036–5046.
- Schillmiller, A. L., Miner, D. P., Larson, M., McDowell, E., Gang, D.R., Wilkerson, C., and Last, R. L. 2010. Studies of a Biochemical Factory: Tomato Trichome Deep Expressed Sequence Tag Sequencing and Proteomics. *Plant Physiol* 153:1212-1223.
- Schmidt, M. A., Hymowitz, T., and Herman, E. M. 2015. Breeding and characterization of soybean Triple Null; a stack of recessive alleles of Kunitz Trypsin Inhibitor, Soybean Agglutinin, and P34 allergen nulls. *Plant Breeding* 134:310-315.
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., et al. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183.
- Schoener, C. S., and Fehr, W. R. 1979. Utilization of Plant Introductions in Soybean Breeding Populations. *Crop Sci.* 19:185–188.
- Sebastian, S. A., and Chaleff, R. S. 1987. Soybean Mutants with Increased Tolerance for Sulfonylurea Herbicides. *Crop Sci.* 27:948–952.
- Sebastian, S. A., Fader, G. M., Ulrich, J.F., Forney, D. R., and Chaleff, R. S. 1989. Semidominant Soybean Mutation for Resistance to Sulfonylurea Herbicides. *Crop Sci.* 29:1403–1408.

- Sebolt, A. M., Shoemaker, R. C., and Diers, B. W. 2000. Analysis of a Quantitative Trait Locus Allele from Wild Soybean That Increases Seed Protein Concentration in Soybean. *Crop Sci.* 40:1438–1444.
- Severin, A.J., Woody, J. L., Bolon, Y. T., Joseph, B., Diers, B. W., et al., 2010. RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol.* 10: 160.
- Shaw, R.H., and Weber, C. R. 1967. Effects of canopy arrangements on light interception and yield of soybeans. *Agron. J.* 59:155-159.
- Singh, B. B., Hadley, H. H., and Bernard, R. L. 1971. Morphology of Pubescence in Soybeans and Its Relationship to Plant Vigor¹. *Crop Sci.* 11:13–16.
- Singh, R. J., and Nelson, R. L. 2015. Intersubgeneric hybridization between Glycine max and G. tomentella: production of F1, amphidiploid, BC1, BC2, BC3, and fertile soybean plants. *Theor. Appl. Genet.* 128:1117–1136.
- Smith, L. G., Hake, S., and Sylvester, A. W. 1996. The *tangled1* mutation alters cell division orientations throughout maize leaf development without altering leaf shape. *Development* 122: 481–489.
- Sneath, P. H. A., and Sokal, R. R. 1973. Numerical Taxonomy. W. H. Freeman, San Francisco.
- Soldatova, O., Apchelimov, A., Radukina, N., Ezhova, T., Shestakov, S., et al., 2005. An Arabidopsis mutant that is resistant to the protoporphyrinogen oxidase inhibitor acifluorfen shows regulatory changes in tetrapyrrole biosynthesis. *Mol. Genet. Genomics* 273: 311-318.

- Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L., and Cregan, P.B. 2013. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS One* 8:1–12.
- Song, Q., Jenkins, J., Jia, G., Hyten, D.L., Pantalone, V., Jackson, S.A., Schmutz, J., and Cregan, P.B. 2016. Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. *BMC Genomics* 17:33.
- Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L., and Cregan, P.B. 2015. Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. *G3 (Bethesda)* 5:1999–2006.
- Stewart, R.T., and Wentz, J.B. 1926. A Recessive Glabrous Character in Soybeans. *Agron. J.* 18:997–1009.
- Stewart, C. N., Adang, M. J., All, J. N., Boerma, H. R., Cardineau, G., Tucker, D., and Parrott, W. A. 1996. Genetic transformation, recovery, and characterization of fertile soybean transgenic for a synthetic *Bacillus thuringiensis* cryIAC gene. *Plant Physiol.* 112:121–129.
- Stijšić, D., Luzzi, B. M., Ablett, G. R., and Tanner, J. W. 1998. Inheritance of Low Linolenic Acid Level in the Soybean Line RG10. *Crop Sci.* 38:1441–1444.
- Stupar, R. M., and Specht, J. E. 2013. Insights from the soybean (*Glycine max* and *Glycine soja*) genome: past, present, and future. *Adv. Agron.* 118:177–204.
- Suhre, J.J., Weidenbenner, N.H., Rowntree, S.C. et al. 2014. Soybean yield partitioning changes revealed by genetic gain and seeding rate interactions. *Agron J.*

106:1631–1642. doi: 10.2134/agronj14.0003

- Szklarczyk, R., M. A. Huyne, and Snel, B. 2008. Complex fate of paralogs. *BMC Evol. Biol.* 8: 337.
- Szymanski, D.B., Marks, M.D., and Wick, S.M. 1999. Organized F-actin is essential for normal trichome morphogenesis in Arabidopsis. *Plant Cell* 11:2331–2347.
- Tadege, M., Ratet, P., and Mysore, K. S. 2005. Insertional mutagenesis: A Swiss Army knife for functional genomics of *Medicago truncatula*. *Trends Plant Sci.* 10:229–235.
- Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Endre, G., Zhao, P. X., Chabaud, M., et al. 2008. Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. *Plant J.* 54:335–347.
- Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.
- Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., et al. 2013. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J* 74:174–183.
- Takagi, Y., and Rahman, S. M. 1995. Variation of different fatty acids in mutants in comparison with natural soybean varieties. *Bull. Fac. Agric. Saga Univ.* 79:23–27.
- Takagi, Y., Hossain, A. B. M. M., Yanagita, T., Matsueda, T., and Murayama, A. 1990.

- Linolenic Acid Content in Soybean Improved by X-Ray Irradiation. *Agric. Bioi. Chem.* 54:1735–1738.
- Takagi, Y., Hossain, M. A. B. M., Yanagita, T., and Kusaba, S. 1989. High linolenic acid mutant in soybean induced by X-ray irradiation. *Jpn. J. Breed.* 39:403–409.
- Takagi, Y., Rahman, S. M., Joo, H., and Kawakita, T. 1995. Reduced and Elevated Palmitic Acid Mutants in Soybean Developed by X-Ray Irradiation. *Biosci. Biotech. Bioch.* 59:1778–1779.
- Takahashi, K., Banba, H., Kikuchi, A., Ito, M., and Nakamura, S. 1994. An Induced Mutant Line Lacking the. ALPHA.-subunit of. BETA.-conglycinin in Soybean (*Glycine max* (L.) Merrill). *Breeding Sci.* 44:65–66.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, Nei, M., et al., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
- Tardivel, A., Sonah, H., Belzile, F., and O'Donoghue, L. S. 2014. Rapid Identification of Alleles at the Soybean Maturity Gene E3 using genotyping by Sequencing and a Haplotype-Based Approach. *Plant Genome*
doi:10.3835/plantgenome2013.10.0034.
- Tuinstra, M. R., Ejeta, G., and Goldsbrough, P. B. 1997. Heterogeneous inbred family (HIF) analysis: A method for developing near-isogenic lines that differ at quantitative trait loci. *Theor. Appl. Genet.* 95:1005–1011.
- Thompson, J. A., and Nelson, R. L. 1998. Utilization of diverse germplasm for soybean yield improvement. *Crop Sci.* 38:1362–1368.

- Tian, Z., X. Wang, R. Lee, Y. Li, Specht, J. E., et al., 2010. Artificial selection for determinate growth habit in soybean. *Proc. Natl. Acad. Sci. U.S.A.* 107: 8563-8568.
- Ting, C.L. 1946. Genetic Studies on the Wild and Cultivated Soybeans. *Agron J* 38:381-393.
- Trick, H. N., Dinkins, R. D., Santarém, E. R., Samoyolov, R. D. V, Meurer, C., Walker, D., Parrott, W. A., Finer, J. J., and Collins, G. B. 1997. Recent advances in soybean transformation. *Plant Tissue Cult. Biotechnol.* 3:9-26.
- Tsuda, M., Kaga, A., Anai, T., Shimizu, T., Sayama, T., Takagi, K., Machita, K., Watanabe, S., Nishimura, M., Yamada, N., et al. 2015. Construction of a high-density mutant library in soybean and development of a mutant retrieval method using amplicon sequencing. *BMC Genomics* 16:1014.
- Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. 2013 From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr Protoc Bioinformatics* 11:11.10.1-11.10.33.
- Vello, N. A., Fehr, W. R., and Bahrenfus, J. B. 1984. Genetic Variability and Agronomic Performance of Soybean Populations Developed from Plant Introductions. *Crop Sci.* 24:511-514.
- Vermeer, J., and Peterson, R.L. 1979. Glandular trichomes on the inflorescence of *Chrysanthemum-morifolium* cv. Dramatic (Compositae). I. Development and morphology. *Can. J. Bot.* 57:705-713.

- Vincent, J. A., Stacey, M., Stacey, G., and Bilyeu, K. D. 2015. Phytic Acid and Inorganic Phosphate Composition in Soybean Lines with Independent IPK1 Mutations. *Plant Genome* doi:10.3835/plantgenome2014.10.0077.
- Voytas, D. F., and Gao, C. 2014. Precision Genome Engineering and Agriculture: Opportunities and Regulatory Challenges. *PLoS Biol.* 12:1–6.
- Walker, A.R., Davison, P.A., Bolognesi-Winfield, A.C., James, C.M., Srinivasan, N., Blundell, T.L., Esch, J.J., Marks, M.D., and Gray, J.C. 1999. The TRANSPARENT TESTA GLABRA1 Locus, Which Regulates Trichome Differentiation and Anthocyanin Biosynthesis in Arabidopsis, Encodes a WD40 Repeat Protein. *Plant Cell* 11:1337–1349.
- Wang, D., Arelli, P. R., Shoemaker, R. C., and Diers, B. W. 2001. Loci underlying resistance to Race 3 of soybean cyst nematode in *Glycine soja* plant introduction 468916. *Theor. Appl. Genet.* 103:561–566.
- Watanabe, S., Hideshima, R., Xia, Z., Tsubokura, Y., Sato, S., Nakamoto, Y., Yamanaka, N., Takahashi, R., Ishimoto, M., Anai, T., et al. 2009. Map-Based Cloning of the Gene Associated With the Soybean Maturity Locus E3. *Genetics* 182:1251–1262.
- Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., Takahashi, R., Anai, T., Tabata, S., Kitamura, K., et al. 2011. A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics* 188: 395–407.
- Wax, L. M., Nave, W. R., and Cooper, R. L. 1977. Weed control in narrow and wide-row soybeans. *Weed Science* 25: 73-78.

- Weber, C. R., and Weiss, M. G. 1959. Chlorophyll mutant in soybeans provides teaching aid. *J. Hered.* 50: 53-54.
- Weber, C. R., Shibles, R. M., and Byth, D. E. 1966. Effect of Plant Population and Row Spacing on Soybean Development and Production. *Agron J* 58: 99
- Weigelt, H. D., Palmer, R. G., and Groose, R. W. 1990. Origin of the *w4-m* allele. *Soybean Genet. Newsl.* 17:81–84.
- Weiss, M.G., and Stevenson, T.M. 1955. Registration of soybean varieties, V. *Agron. J.* 47:541–543.
- Wellburn, A. R., 1994. The spectral determination of chlorophyll a and chlorophyll b, as well as total carotenoids, using various solvents with spectrophotometers of different resolution. *J. Plant Physiol.* 144: 307-313.
- Wilcox, J. R., Cavins, J. F., and Nielsen, N. C. 1984. Genetic alteration of soybean oil composition by a chemical mutagen. *J. Am. Oil Chem. Soc.* 61:97–100.
- Wilcox, J. R., Premachandra, G. S., Young, K. A., and Raboy, V. 2000. Isolation of High Seed Inorganic P, Low-Phytate Soybean Mutants. *Crop Sci.* 40:1601–1605.
- Williams, L.F. 1950. Structure and genetic characteristics of the soybean. In KS Markley, ed, *Soybeans Soybean Prod.* Interscience Publishers, New York, pp 111–134.
- Winter, S. M. J., Shelp, B. J., Anderson, T. R., Welacky, T. W., and Rajcan, I. 2007. QTL associated with horizontal resistance to soybean cyst nematode in *Glycine soja* PI464925B. *Theor. Appl. Genet.* 114:461–472.
- Woodworth, C.M. 1921. Inheritance of cotyledon, seed-coat, hilum and pubescence colors in soy-beans. *Genetics* 6:487–553.

- Wright, S. I., Bi, I., Schroeder, S., and Yamasaki, M. 2005. The Effects of Artificial Selection on the Maize Genome. *Science* 308:1310-1314.
- Xia, Z., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., Anai, T., Sato, S., Yamazaki, T., Lü, S., et al. 2012. Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. *Proc. Natl. Acad. Sci. U.S.A.* 109: E2155–E2164.
- Xu, M., Brar, H. K., Grosic, S., Palmer, R. G., and Bhattacharyya, M. K. 2010. Excision of an active CACTA-like transposable element from DFR2 causes variegated flowers in soybean [*Glycine max* (L.) Merr.]. *Genetics* 184:53-63.
- Yan, A., Pan, J., An, L., Gan, Y., and Feng, H. 2012. The response of trichome mutants to enhanced ultraviolet-B radiation in *Arabidopsis thaliana*. *J. Photochem. Photobiol. B: Biology* 113:29-35.
- Yang, Z., Huang, D., Tang, W., Zheng, Y., Liang, K., Cutler, A.J., and Wu, W. 2013. Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. *PLoS One* 8:e68433.
- Yang, G., Zhang, F., Hancock, C. N., and Wessler, S. R. 2007. Transposition of the rice miniature inverted repeat transposable element *mPing* in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 104:10962–10967.
- Yelverton, F. H., and Coble, H. D. 1991. Narrow row spacing and canopy formation reduces weed resurgence in soybeans (*Glycine max*). *Weed Technology* 5:169-174.
- You, M., Zhao, T., Gai, J., and Yen, Y. 1998. Genetic analysis of short petiole and

- abnormal pulvinus in soybean. *Euphytica* 102: 329-333.
- Yuan, F. J., Zhao, H. J., Ren, X. L., Zhu, S. L., Fu, X. J., and 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.
- Zabala, G., and Vodkin, L. 2003. Cloning of the pleiotropic T locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3' hydroxylase. *Genetics* 163:295–309.
- Zabala, G., and Vodkin, L. O. 2007. A rearrangement resulting in small tandem repeats in the F3 ' 5 ' H gene of white flower genotypes is associated with the soybean W1 locus. *Crop Sci.* 47: S113-S124.
- Zeng, P., Vadnais, D. A., Zhang, Z., and Polacco, J. C. 2004. Refined glufosinate selection in *Agrobacterium*-mediated transformation of soybean [*Glycine max* (L.) Merrill]. *Plant Cell Rep.* 22:478–482.
- Zhang, X.C., Millet, Y., Ausubel, F.M., and Borowsky, M. (2014) Next-Gen Sequencing-Based Mapping and Identification of Ethyl Methanesulfonate-Induced Mutations in *Arabidopsis thaliana*. *Curr. Protoc. Mol. Biol.* 108:7.18.1-7.18.16.
- Zhang, H., D. Zhang, S. Han, X. Zhang, and Yu, D. 2011 Identification and gene mapping of a soybean chlorophyll-deficient mutant. *Plant Breeding* 130: 133-138.
- Zhang, H., J. Li, J. H. Yoo, S. C. Yoo, S. H. Cho et al., 2006. Rice Chlorina-1 and Chlorina-9 encode ChlD and ChII subunits of Mg-chelatase, a key enzyme for chlorophyll synthesis and chloroplast development. *Plant Mol. Biol.* 62: 325-337.
- Zhao, S., Zheng, F., He, W., Wu, H., Pan, S., and Lam, H. M. 2015. Impacts of

- nucleotide fixation during soybean domestication and improvement. *BMC Plant Biol.* 15:1–12.
- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., et al. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33:408–414.
- Zhu, B. G., and Sun, Y. R. 2006. Inheritance of the four-seeded-pod trait in a soybean mutant and marker-assisted selection for this trait. *Plant Breeding* 125:405–407.
- Zou, J. J., R. J. Singh, and Hymowitz, T. 2003. Association of the yellow leaf (y10) mutant to soybean chromosome 3. *J. Hered.* 94: 352-355.
- Zuckerandl, E., and Pauling, L. 1965. Evolutionary divergence and convergence in proteins, pp. 97-166 in *Evolving genes and proteins*, edited by V. Bryson and H. J. Vogel. Academic Press, New York.