Soybean Fast Neutron Mutant Analyses

Yung-Tsi Bolon
411 Borlaug Hall
1991 Upper Buford Circle
University of Minnesota
St. Paul, MN 55108-6026
Office: 612-625-9204
Fax: 651-649-5058
hsie0024@umn.edu

Genome Analysis
Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean

Yung-Tsi Bolon1*, William J. Haun2, Wayne W. Xu3, David Grant4, Minviluz G. Stacey5, Rex T. Nelson4, Daniel J. Gerhardt6, Jeffrey A. Jeddeloh6, Gary Stacey5, Gary J. Muehlbauer2, James H. Orf2, Seth L. Naeve2, Robert M. Stupar2, Carroll P. Vance1,2

*Corresponding author
1United States Department of Agriculture-Agricultural Research Service, Plant Science Research Unit, St. Paul, MN 55108, U.S.A.
2Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, U.S.A.
3Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455, U.S.A.
4United States Department of Agriculture-Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA 50011, U.S.A.
5Division of Plant Sciences, University of Missouri, Columbia, MO 65211, U.S.A.
6Roche NimbleGen, Inc, Research and Development, 504 S. Rosa Rd, Madison, WI 53719, U.S.A.
Footnotes

We are grateful for funding from the National Science Foundation Plant Genome Program (Award #0820769). This work was also supported in part by the U.S. Department of Agriculture, Agricultural Research Service Current Research Information System (CRIS no. 3640-21000-024-00D), the Minnesota Soybean Research and Promotion Council, and the United Soybean Board (Project #0288).

Competing Interests Statement:
The Roche NimbleGen authors (DJG and JJ) recognize a competing interest in this publication.

Corresponding author: Yung-Tsi Bolon, e-mail: hsie0024@umn.edu
Mutagenized populations have become indispensable resources for introducing variation and studying gene function in plant genomics research. In this study, fast neutron (FN) radiation was used to induce deletion mutations in the soybean (*Glycine max* (L.) Merrill) genome. Approximately 120,000 soybean seed were exposed to FN radiation doses of up to 32 Gy to develop over 23,000 independent M2 lines. Here, we demonstrate the utility of this population for phenotypic screening and associated genomic characterization of striking and agronomically-important traits. Plant variation was catalogued for seed composition, maturity, morphology, pigmentation, and nodulation traits. Mutants that showed significant increases or decreases in seed protein and oil content across multiple generations and environments were identified. The application of CGH (comparative genomic hybridization) to lesion-induced mutants for deletion mapping was validated on a mid-oleate X-ray mutant M23 with a known *FAD2-IA* gene deletion. Using CGH, a subset of mutants was characterized, revealing deletion regions and candidate genes associated with phenotypes of interest. Exome resequencing and sequencing of PCR products confirmed FN-induced deletions detected by CGH. Beyond characterization of soybean fast neutron mutants, this study demonstrates the utility of CGH, exome sequence capture, and next generation sequencing approaches for analyses of mutant plant genomes. We present this FN mutant soybean population as a valuable public resource for future genetic screens and functional genomics research.
INTRODUCTION

The release of whole genome sequences in crops such as soybean (Schmutz et al., 2010) marks a new era for genomics research in crop species. Soybean (Glycine max (L.) Merrill) is one of the most valued crops for its ability to fix nitrogen and provide seed protein and oil. Resources to study gene function in this important species are needed, and using mutagenesis to develop population resources has long proven to be a key step for identifying gene function in many organisms.

A number of mutagen sources exist for introducing genomic variation. These include chemical, radiation, and transformation-induced mutagenesis of plant genomes (Ostergaard and Yanofsky, 2004; Waugh et al., 2006; Kuromori et al., 2009). Each of these methods results in a signature footprint of structural variation across the genome (Alonso and Ecker, 2006). Fast neutron (FN) radiation is a particularly promising source of mutagenesis due to the potential to create deletions in a wide range of sizes (Li and Zhang, 2002) for gene knockouts and disruptions.

FN radiation has been used to induce mutations for many decades and has been shown to be an effective mutagen in plants (Koornneef et al., 1982). The majority of mutations that result from FN bombardment are DNA deletions that range in size from a few base pairs to several megabases (Li et al., 2001; Men et al., 2002). Precedence exists in many species, including Arabidopsis (Alonso et al., 2003), Medicago truncatula (Oldroyd and Long, 2003), Glycine soja (Searle et al., 2003), barley (Zhang et al., 2006), and Lotus japonicus (Hoffmann et al., 2007) for the use of FN mutagenesis in forward genetic screens. Many phenotype-associated genes have been successfully identified and cloned through such screens (Meinke et al., 2003).

Recent years have seen renewed interest in FN mutagenesis, and the advent of whole genome technologies has increased the capacity for mutant screening. Comparative genomic hybridization (CGH) microarrays and high-throughput next-generation sequencing (NGS) platforms are powerful tools for the analysis of copy number changes, polymorphisms, and structural variation in the genome. CGH compares DNA hybridization intensities from each genomic sample against a set of fixed probes and is capable of detecting regional copy number changes, such as deletions and duplications (Carter, 2007). NGS can be used to map deletions and insertions, polymorphisms, inversions and translocations (Medvedev et al., 2009). Coupling
whole exome sequence capture with high-throughput NGS can target genes for selective resequencing. Previous studies have shown the utility of each of these approaches to identify and locate genomic changes (Sebat et al., 2004; Hodges et al., 2007; Korbel et al., 2007; Choi et al., 2009). Application of these technologies to a FN mutant population brings detection of gene deletions and phenome-based genetic screens to a whole genome level.

In this study, we describe the development of a FN mutant population resource in soybean (Glycine max). This population was used to screen for seed composition, maturity, morphology, pigmentation, and nodulation mutants. Phenotypes observed in the FN mutant population are described, and deletions within mutated soybean genomes of interest are characterized. We show the promise of the FN mutant population as a community resource and demonstrate the utility of CGH and NGS methods for performing informative genome analyses on such a population.

RESULTS

Soybean fast neutron population resource development

Approximately 60,000 soybean seeds of cultivar M92-220 were irradiated with fast neutrons in the first round of mutagenesis. One quarter (15,000) of the seeds were irradiated at each of the following doses: 4, 8, 16, and 32 Gy. In the first season, 20,000 M₁ seeds, 5,000 from each dose, were planted in two locations and harvested by single seed descent. An additional 60,000 seeds were irradiated in a second round of mutagenesis, half at 16 Gy, and half at 32 Gy, for greater representation in the higher doses, and seeds were also harvested by single seed descent. Independent M₂ plants were propagated in three locations and seed was harvested from individual plants. Statistics on dose comparisons for emergence, survival and propagation percentages were compiled (Table I). As expected, seeds exposed to greater FN radiation doses resulted in fewer emerged M₁ plants in the field and fewer individuals that produced viable seed at the end of the season. In the M₂ generation, radiation dose also correlated inversely to the number of mutant lines that produced seed but did not affect plant emergence in the same manner. As the number of loci that are involved in overall plant development and reproduction are far greater than those limited to plant emergence, this result was not surprising. Since mutant
phenotypes in the M2 generation are more likely due to heritable effects, as opposed to mutant
phenotypes in the M1 generation that may be due to physical damage from FN bombardment, M2
plant leaf tissue and phenotypes were collected and cataloged. Corresponding M3 seed was
collected from independent M2 individuals to form the basis of a FN mutant library resource.

Forward screen for seed composition mutants

An immediate function for the FN soybean population was to serve as a forward genetics
resource for seed composition mutants. To this end, 10,000 M3 lines were screened by NIR
(Near InfraRed) spectroscopy for seed protein or oil percentage of total seed composition.
Additional preliminary data on carbohydrates and fatty acid and amino acid composition was
also determined for a subset of samples. The mean value and detected range for seed
composition components were assayed on seed harvested from three different locations
(Supplemental Table S1) and showed significant increases and decreases in seed protein and oil
within the FN mutant population.

A subset of mutant families consistently displayed increased or decreased seed protein or oil
composition in subsequent generations. Eight seed protein and oil mutants, designated as “PO”
mutants, were chosen primarily based on prior repeat performances in year-to-year rankings.
PO1, PO3, and PO8 are mutants that displayed high seed protein phenotypes and ranked fourth,
second, and first, respectively, among mutants in the M4 generation. PO3 also exhibited
chimeric leaf pigmentation. PO2 possessed the second highest seed oil content detected in the
M4 seed. PO4 and PO5 exhibited the lowest seed oil content, where PO4 also showed high seed
raffinose and sucrose (0.82% and 8.29% of total seed carbohydrate, respectively) content. PO6
exhibited the second lowest seed protein content in the M3 generation and the lowest seed protein
content in the M4 generation. PO7 exhibited the highest seed oil content in the M3 and M4
generations. PO3 showed an increase in combined seed protein and oil content compared to
M92-220, and PO7, PO2, and PO8 also showed minor increases in combined seed protein and oil
content.

The percent protein and oil composition of M6 seed from the 2010 season for mutants PO1-PO8
are displayed (Table II). High seed protein and low seed protein as well as low seed oil
phenotypes were confirmed in 2010. However, the selected high seed oil phenotypes were not observed in the 2010 season, and we attribute this variation to growing season conditions. Several seed protein and oil content mutants displayed maturity differences; however, maturity date did not correlate with a fixed trend in high or low seed protein or oil.

**Fast neutron mutants with visual phenotypes**

Visual phenotypes were recorded for the FN mutant population. Over five hundred independent individuals were observed that displayed an abnormal visual phenotype. Altered phenotypes were observed for approximately two percent of the soybean fast neutron mutant population. Visual phenotypes observed in the field that were not attributed to disease were categorized under the areas of morphology, pigmentation, or maturity. Morphological phenotypes accounted for over seventy percent of the observed abnormal phenotypes. Pigmentation abnormalities were noted for over thirty percent of observed phenotypes, and maturity differences were recorded for around fifteen percent of recorded observations. Some of the mutant lines displayed more than one documented phenotype. Most visual phenotypes were observed for mutants derived from seed treated with 16 Gy (51.7%) or 32 Gy (38.8%) FN radiation doses.

A subset of mutants representing a few of the visual phenotypes observed in the FN population are shown in Figure 1. These include a short trichome mutant (Figure 1A), chimeric and yellow pigmentation mutants (Figure 1B and 1C), and a short petiole mutant with crinkled, curled leaves (Figure 1D). A separate screen for root and nodulation mutants was also conducted within the growth chamber using duplicate M2 seed. Among observed phenotypes during the root and nodulation phenotype screen were a non-nodulating mutant (Figure 1F), a robust mutant with early podset (Figure 1G), and a hypernodulating mutant (Figure 1H).

**CGH validation and marker development for a known FAD2 gene deletion in the M23 mutant**

CGH (comparative genomic hybridization) possesses the potential to map deletions in mutant genomes. Accordingly, a custom NimbleGen soybean 700K-feature CGH microarray containing 696,139 unique soybean probes was designed. This design incorporated unique probes that were
spaced approximately every 1100 bp along the reference soybean genome sequence (www.phytozome.org). To validate the CGH method, genomic DNA from the M23 line was analyzed to verify a known ω-6 fatty acid desaturase (FAD2-IA) gene deletion. Increased oleate content in soybean oil improves its oxidative stability, thereby reducing the risk of producing trans-fatty acids from chemical hydrogenation of processed oil. Several soybean cultivars with increased oleate content are available, one of which is the mid-oleate mutant line M23 (Rahman et al., 1994) derived from X-ray mutagenized seed of the Bay soybean cultivar. Genetic studies indicated that the mid-oleate phenotype of M23 is associated with the deletion of FAD2-IA (Alt et al., 2005; Sandhu et al., 2007; Anai et al., 2008).

Given the agronomic importance of M23 in current breeding programs (e.g., (Scherder et al., 2008) for increased seed oleate content, CGH could also be utilized to define the presence of other genetic lesions in the M23 genome, in addition to the previously reported FAD2-IA deletion. Based on the normalized log2 ratio of M23 to control (cultivar Bay) CGH data, a 163.6 kb CNV (Copy Number Variation) event was detected on chromosome 10 where the M23 hybridization signal was approximately four-fold less than Bay, indicating a deletion (Figure 2A). According to the reference soybean genome sequence, twenty annotated genes are predicted within the chromosome 10 deletion (Supplemental Table S2), including FAD2-IA (Glyma10g42470). No CNV was detected in the other characterized soybean ω-6 fatty acid desaturase genes FAD2-1B, FAD2-2A, FAD2-B and FAD2-C (Schlueter et al., 2007; Pham et al., 2010) (data not shown).

To confirm the chromosome 10 deletion, the genomic region encompassing the deletion junction was amplified by PCR. Flanking primers used for amplification were approximately 1.5 kb from the predicted deletion. As expected, no amplification was detected when genomic DNA from wild-type Bay was used as template, whereas a product of ~3.0 kb was obtained with M23 genomic DNA (Figure 2B). Sequencing of the PCR product precisely mapped the deletion to base positions 49369546 to 49533559 of chromosome 10. Except for an extra nucleotide at the deletion junction, no sequence change was detected in the DNA regions immediately flanking the deletion. The PCR primers designed to amplify this region could be used as a molecular marker to select for the mid-oleate phenotype in segregating breeding populations.
Genomic analysis of fast neutron mutants by comparative genomic hybridization

Many aberrant phenotypes were observed in the soybean FN mutant population and found to be heritable. These phenotypes were hypothesized to result from genomic changes within the mutant after exposure to FN radiation. To confirm and characterize mutations at the genomic level, we performed comparative genomic hybridizations (CGH) using the custom NimbleGen soybean 700K-feature CGH microarray. Thirty microarray hybridizations were performed using genomic DNA from a subset of soybean FN mutants. These mutants were categorized under the following classes: eight seed Protein and Oil (PO), seven Late Maturity (LM), three Early Maturity (EM), three Root and Nodule (RN), and nine aboveground Visual Phenotype (VP) mutants. DNA copy number changes were readily detected in these lines. Full chromosome views of the normalized CGH log2 hybridization ratios of mutant versus M92-220 control are shown (Figures 3 and 4, Supplemental Figures S1-S3).

Eight seed protein and oil composition mutants were chosen for CGH analyses based on high or low seed protein or oil composition across multiple environments and generations (Table II). The M4 generation was used for CGH analyses, and high/low phenotypes were confirmed on the M5 seed harvested from an M4 plant. Genomic regions that exhibited changes in DNA copy number were detected in these soybean FN mutants and collectively displayed in Figure 3.

Soybean FN mutants with aboveground visual phenotypes were also assayed by CGH to detect and map genomic changes. VP1 (Figure 1A) and VP2 mutants were independently recovered and both exhibited a short trichome phenotype. VP3-VP5 pigmentation and short petiole mutants (Figure 1B-D) were also assayed. In addition, a copper leaf colored mutant (VP6), a mutant with abnormal floral meristem development (VP7), a mutant showing fused trifoliates (VP8), and a mutant exhibiting thick, twisted petioles (VP9) were subjected to CGH for genomic analysis. DNA copy number changes detected by CGH for these visual phenotype mutant lines are displayed in Figure 4. Also shown are genomic regions with DNA copy number changes detected by CGH for late maturity, early maturity, and root and nodulation mutants (Supplemental Figures S1-S3).

Microarray hybridizations were performed using genomic DNA from M2, M3, or M4 plant tissue. Of the thirty hybridization results, a subset of M4 seed composition and maturity mutant results
were derived from a common M2 or M3 plant. Thus, several detected DNA copy number change regions coincided (Figure 3 and Supplemental Figure S1: Chromosome (Ch) 10, ~48.7 Mb (PO1, LM1, LM5); Ch 10, ~23 Mb (PO4, PO5)) due to shared pedigree. These results refer only to the relationship of specific mutant plants that were assayed by CGH and not to mutant lines in general that are unique.

**Screening for false-positive CNV events in the soybean FN population**

A number of genomic locations consistently exhibited CNV across many FN mutant genotypes. These can be observed in the multicolored groupings above or below the axis (Figures 3 and 4, Supplemental Figures S1-S3), for example, on chromosomes 9, 18 and 20. To further characterize and delineate these regions, we performed SNP genotyping across a subset of the soybean FN mutant population using the Illumina Goldengate SNP genotyping platform with 1536 Universal soybean SNP markers (Hyten et al., 2010). Twenty-four markers detected SNPs within the population on the following chromosomes: 3, 6, 9, 14, 16, 18, and 20 (Supplemental Table S3, Supplemental Figure S4). These data indicate that there were several regions of genomic heterogeneity maintained among the M92-220 individuals within the population (bulked seed) that was exposed to FN mutagenesis. This type of intra-cultivar heterogeneity appears to be typical of many soybean accessions (Haun et al., 2011). Segregation of regions of genomic heterogeneity that include natural copy number polymorphisms may appear as false-positive fast neutron-induced duplications or deletions in CGH CNV analyses.

We used a combination of CGH and SNP genotyping data to mask regions of genomic heterogeneity from FN CNV analysis. Microarray probes located within confirmed polymorphic SNP regions were removed from consideration as FN-induced polymorphisms. The degree of variation at each probe position across 30 CGH microarrays was visualized by calculating and plotting values that crossed the 95th percentile log2 CGH ratio threshold (Supplemental Figure S4). Major peaks of variation and SNP locations coincided to confirm regions of genomic heterogeneity within the population and to differentiate these regions from DNA copy number variation arising from other sources, i.e., FN bombardment.
Genomic CNV events detected in fast neutron mutants

Analysis of CGH data revealed a total of 61 genomic DNA regions with CNV among the 30 lines tested. This number was calculated using a stringent threshold of segments greater or less than three standard deviations from the mean normalized log$_2$ ratio for the sample versus control. Of the 61 CNV events that passed the threshold, 52 were putative deletion regions (85.25%), and 9 were putative duplication regions (14.75%). The average number of regions with CNV per mutant was 2.03, and the average number of deletion regions detected per mutant was 1.73.

Our stringent analyses criteria were designed to identify homozygous deletions. Several heterozygous or hemizygous events were potentially observed. For example, CGH variation detected in RN1 and VP7-9 did not pass the threshold, possibly due to allelic differences in genomic DNA from M$_2$ generation individuals. In VP9, deviations from background and control hybridizations appear (Figure 4: Ch 1, 6, 9, 11) near the threshold. Other putative heterozygous or hemizygous events were observed, for example, in Figure 4: VP4 on Ch09, VP5 on Ch17, and VP7 on Ch20. In comparison, candidate homozygous genomic deletions reach a greater log$_2$ ratio amplitude (Figure 3: Ch10 and 18). A mixture of these two deletion types is observed in PO2 (Figure 3: Ch 2).

Among the 61 genomic regions exhibiting significant CNV, putative deletions and duplications ranged in size from 986 bp to almost 3 Mb, and the mean size of a DNA copy number change region was around 367 kb. The mean and median size of the total detected DNA copy number change regions per mutant was 777 kb and 319 kb, respectively. A summary of the detected CNVs from 30 CGH microarrays for seed composition, maturity, root and nodule, and other visual phenotypes is shown (Supplemental Table S4). Detected CNV regions were found on every chromosome and distributed as follows: Ch01: 3, Ch02: 5, Ch03: 2, Ch04: 1, Ch05: 2, Ch06: 5, Ch07: 2, Ch08: 1, Ch09: 4, Ch10: 8, Ch11: 2, Ch12: 1, Ch13:1, Ch14: 2, Ch15: 2, Ch16: 4, Ch17: 7, Ch18: 2, Ch19: 6, Ch20: 1.

The number of genes located within each of the 61 significant CNV events ranged from zero to 145. On average, seventeen genes were found per CNV event. A total of 1048 genes, including 634 high confidence genes, were found within all CNV events defined across the 30 CGH microarrays (Supplemental Table S5). For 130 of these genes, putative paralogous genes were
found elsewhere in the genome. Over half of the detected CNV events occurred in pericentromeric regions, with 29 of 52 deletions and 5 of 9 duplications. This finding may be expected, as approximately half of the genome space consists of pericentromeric regions. Importantly, deletion regions that contained single genes and mutants with as few as four predicted genes within all detected deletions were recovered.

**Genomic analysis of soybean fast neutron mutants by exome resequencing**

To confirm CGH-detected deletions in FN mutants, exome capture and resequencing was performed on four of the thirty soybean FN mutants assayed by CGH. Mutant DNA libraries were constructed and hybridized to a NimbleGen capture array designed to cover the soybean exome. The design comprises 69% of the Glyma 4.0 annotated coding sequence (CDS) features with total capture space that represents 52.3 Mbp and targets 226,207 CDS features (Haun et al., 2011). DNA captured by the array was amplified and sequenced by high-throughput Illumina NGS technologies to carry out exome resequencing of the mutant DNA.

The exomes of four mutants, PO1, PO8, VP1, and VP5, were resequenced. Exon counts were normalized to the total counts in each sample and visualized to display deleted exons (Figure 5). Single chromosome views for the main deletions detected by CGH are shown for PO1 and PO8 (Figure 5A and C). CGH analyses detected deletions on chromosome 10 at ~48.7 Mb in PO1 and on chromosome 16 at ~28.1 Mb in PO8. Exome resequencing confirmed these deletions, as there was no evidence for the existence of exon sequence in these respective regions in the mutant sample DNA, while the control sample did provide sequence reads that mapped to these regions (Figure 5B and 5D). Similarly, exome resequencing and whole genome paired-end mapping with NGS confirmed a CGH-detected deletion on chromosome 13 at ~42.3 Mb in VP5 (data not shown). Exome resequencing was performed on genomic DNA from a sibling of the short trichome VP1 mutant assayed by CGH; this sibling also displayed the short trichome phenotype. In addition, whole genome paired-end mapping using NGS was performed on VP1 (data not shown). The same deletion was detected on chromosome 5 at ~36.4 Mb by three different approaches, CGH, exome resequencing, or whole genome paired-end mapping through NGS, in two individuals with the VP1 short trichome mutation.
Demarcation of fast neutron deletion regions and cosegregation of a deletion with a dominant mutant phenotype

A small deletion detected by both CGH and exome resequencing in VP1 contained a single gene (Glyma05g31280) encoding a tetratricopeptide repeat-containing chaperone binding protein. This region was chosen for polymerase chain reaction (PCR) confirmation. Primers were designed within the CGH probe sequences that flanked the deletion region, and PCR was performed under short extension conditions on VP1 versus wild-type M92-220 genomic DNA templates. Gel electrophoresis of the PCR products was performed. A single 579 bp product was obtained from VP1 and not seen in the wild-type control (Figure 6A). Upon sequencing of the 579 bp product and alignment to the reference genome sequence, the exact breakpoint sites for the deletion were determined to span chromosome 5 base positions 36426532-36430207 (Figure 6B). A larger deletion of nearly 40 kb on chromosome 10 that was detected by CGH (Figure 5A) and exome resequencing (Figure 5B) was also confirmed by PCR in PO1 (Figure 6C and D).

A putative heterozygous deletion on chromosome 17 detected by CGH and by whole genome paired end mapping through NGS (data not shown) in VP5 was also confirmed by PCR. Sequencing of the PCR product (Figure 6E) revealed an 837,919 bp deletion on chromosome 17. This deletion spans chromosome 17 base positions 7770585-6932666, encompasses 87 high-confidence genes, and interrupts a ubiquitin-specific proteinase gene at one end. M₃ progeny of VP5 segregated ~3:1 for the short petiole phenotype, indicative of a dominant mutant phenotype. The chromosome 17 deletion was found to cosegregate with the mutant phenotype (Figure 6E and F) in all fourteen M₃ progeny. We used CGH to genotype six of the M₃ progeny (data not shown) and found three mutant individuals heterozygous for the chromosome 17 deletion. A total of 41 M₄ individuals derived from the heterozygous M₃ parents were scored for plant architecture and the chromosome 17 deletion. A perfect correlation was found between the presence of at least one copy of the chromosome 17 deletion and the short petiole phenotype among the 30 mutant and 11 wild-type segregating individuals. These data suggest that the chromosome 17 hemizygous deletion may be sufficient to confer the mutant phenotype in a dominant fashion or is tightly linked to the dominant causative locus. Furthermore, these data
provide evidence for the potential to detect associated loci and develop specific markers for mutant phenotypes within this soybean FN mutant population.

Soybean fast neutron mutant database

The complete catalog of soybean fast neutron M2 mutants from this study has been launched at http://www.soybase.org/mutants. FN mutant trait data, observed phenotype descriptors, and photographs are presented on this site along with parallel data from the unmutagenized wild-type M92-220. Currently, the soybean FN mutant database lists over 23,000 independent FN mutant lines. The original seed composition data on M3 seed from over ten thousand independent mutants is also displayed on the site.

Information compiled on the soybean FN mutant population is available for users to browse and search for recorded phenotypes of interest. A menu with key descriptors is provided to facilitate searches, and recorded trait ranges across the mutant population are displayed upon trait selection to assist in data filtering. Photographs of mutant plants and plant parts are presented in a user-friendly browser. Images of special interest mutants chosen for further analysis may be accumulated during browsing to view multiple images on a single page, facilitating comparisons between mutants. Standardized soybean trait ontology tags are associated with each descriptor or measured feature. Seed availability is noted for each mutant line. Over 240 mutants with observed visual phenotypes have available seed. Bulk M3 and M4 seed from the soybean FN mutant population are also available.

FN mutant genomic analyses results, observations connected to the original M2 identifier and seed status are continually updated in this dynamic database. Multi-chromosome diagrams and normalized log2 hybridization ratios for each CGH microarray experiment are available for viewing and download at the soybean FN population database. Sequence homology search capabilities through BLAST (Altschul et al., 1990) allow the user to find mutants with CNV events that cover a gene with nucleotide or protein sequence similarity to a sequence of interest. An added “mutant” track for FN mutants on the genome browser displays the location of all CNV events defined in this study and leads to information on the genes within identified deletion and duplication regions.
DISCUSSION

In this study, we established a soybean FN mutant population and performed phenotypic and genomic analysis on mutants of phenotypic interest. We identified soybean FN mutants with seed composition phenotypes and mapped gene deletions within these mutants. Additional phenome analyses were performed by selecting mutants with observed visual phenotypes and identifying genomic CNV events. Notably, this study combines complementary genome-wide microarray-based and NGS technologies to map FN-induced deletions. This population is a resource for studying gene function, and the soybean FN mutant database provides open access to this resource.

Existing mutant resources

A number of resources currently exist for phenome analysis in plant species. Open databases are available for Arabidopsis (Sundaresan et al., 1995; Martienssen, 1998; Tzafrir et al., 2003; Tzafrir et al., 2004; Kuromori et al., 2006), rice (Zhang et al., 2006; Miyao et al., 2007; Larmande et al., 2008), maize (Fernandes et al., 2004; Lawrence et al., 2007), sorghum (Xin et al., 2008), barley (Caldwell et al., 2004), and tomato (Menda et al., 2004), among others. A number of mutant population resource databases exist for legume species, including Tnt-1 (Tadege et al., 2008) and fast neutron bombardment (Wang et al., 2006; Rogers et al., 2009) populations in Medicago truncatula, EMS TILLING populations in soybean (Cooper et al., 2008) and Lotus japonicus (Kuromori et al., 2009), and now a soybean FN mutant database comprised of the mutants developed in this study. In terms of recovering viable M$_2$ plants with observable mutant phenotypes, the most successful fast neutron dose rates were 16 to 32 Gy for our soybean population. This was comparable to the optimized FN doses of 18-20 Gy reported for rice (Li et al., 2001) and 32.5 Gy reported for Medicago truncatula (Rogers et al., 2009). This study describes the assembly and characterization of the largest collection of soybean FN mutants to date accompanied by the only FN mutant database to display genome-wide coverage of deletion events in addition to recorded phenotypic traits.
Forward genetic screening and prospects for reverse genetic screening

Mutant populations are sources of increased genetic and phenotypic diversity. Three phenotypic categories, physical, chemical, and biological, have been proposed for the application of phenome analysis to mutant resources (Kuromori et al., 2009). In this study, we described visual or physical phenotypes observed in the soybean FN mutant population as well as chemical phenotypes profiled by NIR spectroscopy of mutant seed. Biological conditional phenotypes, involving screens in the presence of different stresses or growth conditions, are also possible. Forward screens for drought resistance, SCN (soybean cyst nematode) resistance, and yield have been initiated in the soybean FN mutant population. Additional seed composition and root and nodulation screens are also in progress.

Fast neutron mutant populations for reverse genetic screening have been created for Arabidopsis (Li et al., 2001), *Medicago truncatula* (Rogers et al., 2009), and rice (Wu et al., 2005). The resources established in this study provide the basis for future reverse genetic screening of genes of interest. A library of DNA isolated from unique M2 individuals is under development and will allow for the screening and recovery of mutants with DNA copy number changes at specific locations. Previous estimates for the number of mutants required for saturation mutagenesis of a plant genome were based on estimates calculated through the success rate of screening for single gene mutations (Li et al., 2001; Rogers et al., 2009) and prior estimated correlations with observed albino frequencies (Koornneef et al., 1982; Rogers et al., 2009). Through the use of a genome-wide CGH platform, we obtained a more comprehensive view of the number and size of deletions that are present within our soybean FN mutant lines. In a simplified calculation, if approximately one thousand genes are deleted per thirty mutants, as seen in this study, our collection of over 20,000 mutants may provide 10x coverage of the soybean genome.

Large-scale reverse genetic screening is now feasible with the availability of the soybean genome sequence. PCR-based strategies, such as Deleteagene (Li et al., 2002) and De-TILLING (Rogers et al., 2009), designed to screen fast neutron mutant populations for mutations at individual genes of interest, may be facilitated in the future by screening pools of mutant DNA through paired-end next-generation sequencing to first identify the location and size of potential
deletions within the population. Through paired-end next-generation sequencing of multiplexed mutant genomic DNA samples, we were able to detect deletions confirmed by CGH (data not shown). Once identified by NGS and deposited as putative deletions in a database, deletions of interest may then be chosen for confirmation by PCR. Such a strategy, combined with the pooling methods described previously (Rogers et al., 2009), may facilitate high-throughput reverse genetic screening. With a population size of over 23,000 mutants and the tools launched in this study for mutant genome characterization, the development of a tiled set of identified deletions and duplications within the mutants may be possible. The establishment of such a resource would be of great value for future research in soybean functional genomics.

Genomic mutation detection methods

The Nimblegen 700K soybean CGH microarray was designed with unique probes spaced approximately every 1100 bp along the soybean genome sequence for a total of 696,139 probe sequences. This coverage allowed for preliminary whole genome information to be gained regarding the effect of FN radiation exposure on soybean genomic DNA. The resolution of the microarray platform and analysis method was limited in many regions to approximately 2 kb due to the requirement for a CNV event to cover at least two adjoining probes. In reality, we were able to detect a 986 bp deletion because the spacing between probes in this case was closer than average at ~450 bp. It is also possible that a single deviant datapoint on the CGH array could represent a true deletion, however such a region would not be identified using our stringent filtering criteria, and the exact breakpoints for fragmented deletion borders may not be resolved using CGH alone. The size of FN-induced deletions defined in this study range from less than 1 kb to almost 3 Mb and expands the range of FN-induced deletions reported in plants (Li et al., 2001; Men et al., 2002). However, an underestimation of structural variation events, particularly of smaller deletions and duplications, is likely to occur.

To validate the use of CGH for detecting gene deletions, we performed CGH on the mid-oleate X-ray mutant M23 with a known \textit{FAD2-1} gene deletion. Through CGH, we delineated an X-ray-induced genomic deletion to an estimated size of 163.6 kb on chromosome 10, encompassing the \textit{FAD2-1A} gene, and subsequent sequencing of the region showed that the deletion covered
164.01 kb. This control provided a proof of concept that CGH analysis can be reliably used to identify and develop markers for important deletion breakpoints in the soybean genome.

The use of paired-end mapping NGS technologies to detect structural variations in the genome has been reported previously (Korbel et al., 2007). In this study, we combined paired-end mapping NGS with array capture technologies to detect the presence and absence of gene exons in mutant versus control genomic DNA. Our use of NGS combined with exome capture and whole genome paired-end mapping confirmed CGH-detected deletions. Hybridization-based platforms are subject to many sources of error that may reduce the power of detecting polymorphisms, particularly for small features. The addition of NGS approaches allowed us to impose stringent parameters that, in conjunction with CGH, resulted in high-confidence calls, particularly for polymorphic genomic regions. The use of NGS paired-end technologies also adds the utility of mapping the insertion location of genomic duplications and translocations. As NGS costs decrease in the future, the cost per genotype will likely become comparable or preferable to CGH and enable the high-throughput genotyping of lesion-induced mutant populations.

Connecting gene to function

Genes detected in copy number change regions are candidate genes for the associated traits of interest. Assessment of existing annotations reveals potential genes that may play a role in the observed phenotype. For example, genes involved in regulation of protein metabolism (Glyma10g18310, encoding ubiquitin-conjugating enzyme) and proteolysis (Glyma10g19260, encoding serine carboxypeptidase) are deleted in low seed oil mutants (PO4 and PO5, Supplemental Table S5). Transcripts for a gene encoding a ubiquitin-conjugating enzyme were previously observed to accumulate during seed development of a high seed oil soybean cultivar (Wei et al., 2008). In addition, a gene involved in protein biosynthesis (Glyma02g03750, encoding threonyl-tRNA synthetase) is absent in a high seed oil mutant (PO2). Examination of differentially accumulated transcripts in soybean near-isogenic lines also showed higher accumulation of a gene encoding a tRNA synthetase in the higher seed oil line LoPro (Bolon et al., 2010). Furthermore, a gene involved in lipid metabolism (Glyma10g26510, encoding lipase
(class 3)) is deleted in a low seed protein mutant (PO6). These observations may reflect changes in flux and relate to the inverse correlation between protein and oil levels observed in the soybean seed (Bolon et al., 2010).

Likewise, a detected deletion region in late maturity mutant LM6 contains genes (Glyma04g36620 and Glyma04g36630) encoding jumonji domain transcription factors with homology to Arabidopsis REF6 (Relative of Early Flowering 6), a suppressor of Flowering Locus C expression that delays flowering (Noh et al., 2004). Deletion of the REF6 gene may explain or contribute to the late maturity phenotype observed in LM6. At least 15 genes involved in the regulation of transcription were found in deletion regions identified in this study. These include AP2, Dof, WRKY, helix-loop-helix, NAM, bZIP, and B3 domain transcription factors (Supplemental Table S5). Previously identified QTL regions for seed protein, seed oil, and pod maturity exist on chromosomes with detected deletions. Additional mapping and follow-up studies will be required to determine whether these regions coincide.

Greater evidence for the existence of functional genes within FN mutant deletion regions may be found through the use of complementary genome resources. For example, examination of genes within deletion regions that possess minimum transcript accumulation evidence of at least ten read counts in at least one tissue of the recently reported RNA-Seq soybean gene expression atlas (Libault et al., 2010; Severin et al., 2010) reduces the gene list from 1048 to just over two hundred genes. Of these genes, only ~150 do not possess paralogs elsewhere in the genome. Using such deduction methods, the potential pool of candidate genes may be condensed. The genome browser function at SoyBase (www.soybase.org/gbrowse/cgi-bin/gbrowse (Grant et al., 2009)) allows for direct comparison of the location of soybean fast neutron deletions and duplications detected in this study to soybean gene models, gene expression evidence, and duplicated gene segments.

The assembly of a FN mutant population in soybean facilitates the study of genes with loss-of function phenotypes. Our findings support the utility of FN radiation as a mutagen to delete gene regions with tandem duplications. For example, eighteen genes with F-box and WD domain annotations (Glyma15g19120-Glyma15g19290) were found within a deletion region detected in EM3 (Supplementary Table S5). Multiple leucine-rich repeat genes (Glyma16g27520-Glyma16g27560), aminotransferase-related genes (Glyma09g07320-
Glyma09g07340) and pentatricopeptide repeat genes (Glyma16g27780-Glyma16g27800) are tandemly arrayed within deletion regions detected in EM1. The use of FN bombardment to delete genes facilitates the identification of genes that are not functionally redundant. Additional genomic and cytogenetic studies involving this population may also provide insight into the minimal necessary soybean genome.

Further analyses are required to fully characterize the phenotypes identified in this study and to confirm gene candidates. Map-based cloning and confirmation of candidate gene function by complementation have successfully been performed using FN legume mutants to characterize nodulation genes NSP2 (Oldroyd and Long, 2003; Searle et al., 2003; Kalo et al., 2005) and DMI1 (Ane et al., 2004) in Medicago truncatula and NTS-1 (Men et al., 2002) in Glycine soja. Microarray-based strategies together with bulked segregant analysis (Gong et al., 2004) or transcript profiling (Mitra et al., 2004) have also resulted in gene cloning from deletion mutants. Together with NGS technological advances, the utility of a fast neutron mutant population may now be explored more rapidly and comprehensively than ever.

MATERIALS AND METHODS

Resource development

Soybean seeds of mid-maturity group I cultivar M92-220, derived from the 2006 Crop Improvement Association seed stock of the variety MN1302 (Orf and Denny, 2004), were exposed to fast neutron radiation doses of 4, 8, 16, and 32 Gy at the McClellan Nuclear Radiation Center at UC-Davis. M1 seed was planted and propagated by single seed descent. M2 seed was planted in a grid format, and M2 plants were individually tagged with an assigned barcode identifier. Young leaf tissue was collected for each M2 plant, and M3 seed was single-plant harvested from greater than 20,000 M2 plants. Seeds were analyzed for seed composition by NIR using a Perten DA7200 diode array instrument (Huddinge, Sweden) equipped with calibration equations developed by Perten in cooperation with the University of Minnesota.
SNP genotyping

Genomic DNA samples from soybean leaf tissue were isolated using the Qiagen DNeasy kit protocol. DNA samples were then assayed on the Illumina Goldengate® platform for genotyping using 1536 soybean Universal SNP BARC markers (Hyten et al., 2010). Polymorphic SNP markers across the soybean FN population were detected, and SNP markers were subject to BLAST (Altschul et al., 1990) analysis to recover the physical SNP position along the reference genome sequence.

Comparative Genomic Hybridization (CGH)

Comparative genomic hybridizations were performed as previously described (Haun et al., 2011) on the NimbleGen Glycine max CGH microarray, which consists of 696,139 unique oligonucleotide probes (50-70mers) designed from the reference Williams 82 sequence (Schmutz et al., 2010) and spaced at ~1.1 kb intervals (platform details can be found in GEO accession number GPL11198 at http://www.ncbi.nlm.nih.gov/geo/). Mutant (Cy3 dye) and reference (Cy5 dye) labeling reactions were performed with 1μg each of genomic DNA from mutant and pooled mutant leaf tissue samples. Pooled reference samples consisted of DNA from 35 independent M92-220 mutant families. The same reference pool was used as the Cy5 dye in all hybridizations. WT (wild-type) M92-220 and WT control pool samples were also labeled for comparison purposes against the pooled mutant reference. For the M23 control, M23 X-ray mutant DNA was labeled with Cy3 dye and Bay wild-type DNA was labeled with Cy5 dye. Labeled DNA was quantified and then hybridized for 72 hours at 42°C on the CGH microarrays. PO1, LM1, and LM5 are related. PO6 and LM7 were M4 samples derived from the same M3 plant. PO4 and PO5 were M4 samples also derived from a common M3 individual. Supplemental Table S6 provides the database and pedigree key for the 30 FN mutant lines that were analyzed by CGH.

CGH analysis of CNV events
CGH data was analyzed using the Roche NimbleGen NimbleScan v2.5 segMNT algorithm. Corrected log2 ratios were obtained for each probe datapoint. Probe segmentation and corrected log2 ratios were obtained for each probe datapoint essentially as described (Haun et al., 2011). Parameters were set for minimum segment lengths of two probes and segment log2 ratio differences of 0.1 between segments at the 0.999 acceptance percentile. Spatial correction and qspline normalization was applied. Significant copy number changes were determined by retrieving segments with an average corrected log2 ratio greater than the average plus three standard deviations (increase) or less than the average minus three standard deviations (decrease). If a gap between potential segments was less than half the size of the total distance covered by neighboring segments, then the entire region was considered a single CNV event. Final mutant deletion and addition regions were determined after filtering each CNV event against regions of natural heterogeneity determined through SNP genotyping and CGH analysis of mutants within the population. Genes (ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v4.1/Gmax/annotation/Glyma1.gff2) that overlapped CNV events were determined using a custom Perl script. Paralogous genes in soybean were identified using BLAST (Altschul et al., 1990), DAGChainer (Haas et al., 2004), and selection of gene pairs from synteny blocks with average Ks values between 0.03 and 0.60.

**CGH analysis of natural genetic heterogeneity within the mutant population**

The standard deviation from the average corrected log2 ratio of Cy3 (sample) to Cy5 (control) intensities was calculated at each probe position for each CGH array. After the average absolute value of the above was calculated for each probe position across thirty CGH arrays, the 95th percentile border (1.103589) was calculated across 696,139 unique probe positions, and the median value across each eleven-probe sliding window was determined. Regions with median values that peaked above the 95th percentile border were candidate regions of genetic heterogeneity highlighted for further examination.

**Exome resequencing and data analysis**
Genomic DNA preparations from four mutants, PO1, PO8, VP1, and VP5, were extracted using the Qiagen (Valencia, CA) Plant DNeasy system. Library preparation, exome capture, amplification, and high-throughput sequencing of exome-captured libraries were performed as previously described (Haun et al., 2011). The Illumina Solexa 76-base paired end short read sequences were aligned to the soybean genome sequence version 4.1 (Gmax.main_genome.scaffolds assembly; ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v4.1/Gmax/assembly/sequences/) using software SOAP2. The unique alignment allowed for a maximum of two mismatches. The Glyma v5.0 annotation file Glyma1_highConfidence.gff3 (2/8/2010) was used for exon annotations. There are 55,787 mRNAs and a total of 345,213 exons in this high confidence annotation. After unique alignment of paired read sequences to the reference soybean genome, the number of reads in any direction at each exon in each gene was counted using a custom Perl script. For read counts, a minimum of 70 out of the 76 bases of read sequence was required to overlap the reference exon sequence. The count number was globally normalized by dividing each number by the total counts in the sample. For visualization of exome differences, one count was added to each value, and the log2 ratio of the normalized sample over control count number was calculated and plotted.

**PCR analysis**

Select regions with detected deletions were chosen for confirmation by PCR. Primers were designed within CGH probes that flanked the detected deletion region for a PCR product that spanned the deletion site. The PCR product was gel-extracted and sequenced. Alignment of the PCR product sequence to the reference *G. max* genome sequence assembly (www.phytozome.net) was performed to determine the exact breakpoint borders. The following primers were used for PCR confirmation: VP1 5’-GTA AGT AGC CTA CGC ATG ACC-3’ (forward) and 5’-CAA TGT GAC CAA GCA CTG ACA C-3’ (reverse), PO1 5’- CAC TTT CCG GTA AGA TTA AGG G -3’ (forward) and 5’-CAG TTT GCT TAC ACT CTG ACT C-3’ (reverse), VP5 5’- TAT AAA GAG GGA AGG TTT GTG C -3’ (forward) and 5’-CAT GGG CAA ACT ATT ATG CTT G -3’ (reverse), M23 5’- CCA CAT CCT GAA TAT TCG GAA TCT GTG AA-3’ (forward) and 5’-GTG AAG CAA CAT ACC TTG ATG TCT TCG AT-3’ (reverse).
ACKNOWLEDGMENTS

We thank Jeffrey Roessler, Dimitri Von Ruckert, Renee Schirmer, and Gabriel Bascur Bascur for technical and field support, Dr. Zheng Jin Tu for supercomputing support, and Mauricio Assuncao, Phil Schaus, Arthur Killam, Dhananjay Mani, Tracy O’Neil, Dr. Jill Miller-Garvin and numerous students for their contributions to this project. In addition, we thank Dr. Bruna Bucciarelli for photography and plant ontology support, Dr. David Hyten for Goldengate SNP genotyping, Dr. Steven Cannon for soybean paralog data, and Dr. Kristin Bilyeu for the M23 line. Special thanks to Kevin Feeley and Nathan Weeks for database and website support.

We would like to acknowledge the use of resources at the MSI Supercomputing Institute at the University of Minnesota. Mention of trade names or commercial products in this report is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.
LITERATURE CITED


reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11: 38


Pham AT, Lee JD, Shannon JG, Bilyeu KD (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


Sandhu D, Alt JL, Scherder CW, Fehr WR, Bhattacharyya MK (2007) Enhanced oleic acid content in the soybean mutant M23 with the deletion in the Fad2-1a gene encoding a fatty acid desaturase. J Amer Oil Chem Soc 84: 229-235


FIGURE LEGENDS

**Figure 1: Selected soybean fast neutron mutants with visual phenotypes.** A) A short trichome mutant VP1 (top) displays short trichomes compared to wild-type soybean (bottom). B) A chimeric mutant VP3 with altered leaf pigmentation patterns is visible among a field of other soybean mutants with normal pigmentation. C) A row of petite yellow-tinged mutants (VP4). D) The short petiole and curled, crinkled leaf mutant (VP5). E) Wild-type root and nodules are shown at three weeks post germination. F) Mutant RN1 displays abnormal shoot-root connection and does not appear to nodulate. G) Robust mutant RN2 displays elongated internodes compared to wild-type and later begins pod set precociously. H) Hypernodulating mutant RN3 displays increased nodulation compared to wild-type in (E).

**Figure 2: Detection of a genomic deletion region encompassing a known deleted gene in M23.** A) CGH analysis identifies a deletion at the end of chromosome 10. The Y-axis represents unaveraged log<sub>2</sub> ratios of M23 to Bay hybridization signals. B) Confirmation of the predicted deletion by PCR. Arrows represent flanking amplification primers located 1.5 kb from the predicted 163.6 kb deletion.

**Figure 3: CNV events detected by CGH in soybean fast neutron seed protein and oil mutants.** Full chromosome views of CNV events are depicted for all 20 soybean chromosomes. The normalized log<sub>2</sub> ratio of sample to control data is plotted as the median across 11 probe datapoints across chromosome positions. Results from each array are color-coded for mutants PO1 through PO8. PO1-PO3, PO8 = high seed protein. PO4 = low seed oil and high seed raffinose and sucrose. PO5 = low seed oil. PO6 = low seed protein. PO7 = high seed oil. The gray overlay represents CGH data from control vs. control comparative genomic hybridizations. Colored regions above and below the control regions potentially represent copy number change differences. The y-axis scale is in terms of the number of standard deviations from average with the segment threshold for deletions or duplications at +/-3.

**Figure 4: CNV events detected by CGH in soybean fast neutron visual phenotype mutants.** Full chromosome views of CNV events are depicted for all 20 soybean chromosomes. The normalized log<sub>2</sub> ratio of sample to control data is plotted as the median across 11 probe datapoints across chromosome positions. Results from each array are color-coded for mutants.
VP1 through VP9. VP1 and VP2 = short trichomes. VP3 = chimeric leaf pigmentation. VP4 = petite and yellow leaf. VP5 = short-petiole and crinkled leaf. VP6 = copper leaf. VP7 = abnormal floral meristem development. VP8 = fused trifoliates. VP9 = thick, twisted petioles. The gray overlay represents CGH data from control vs. control comparative genomic hybridizations. Colored regions above and below the control regions potentially represent copy number change differences. The y-axis scale is in terms of the number of standard deviations from average with the segment threshold for deletions or duplications at +/-3.

Supplemental Figure S1: CNV events detected by CGH in soybean fast neutron late maturity mutants. Full chromosome views of CNV events are depicted for all 20 soybean chromosomes. The normalized log2 ratio of sample to control data is plotted as the median across 11 probe datapoints across chromosome positions. Results from each array are color-coded for mutants LM1 through LM8. The gray overlay represents CGH data from control vs. control comparative genomic hybridizations. Colored regions above and below the control regions potentially represent copy number change differences. The y-axis scale is in terms of the number of standard deviations from average with the segment threshold for deletions or duplications at +/-3.

Supplemental Figure S2: CNV events detected by CGH in soybean fast neutron early maturity mutants. Full chromosome views of CNV events are depicted for all 20 soybean chromosomes. The normalized log2 ratio of sample to control data is plotted as the median across 11 probe datapoints across chromosome positions. Results from each array are color-coded for mutants EM1 through EM3. The gray overlay represents CGH data from control vs. control comparative genomic hybridizations. Colored regions above and below the control regions potentially represent copy number change differences. The y-axis scale is in terms of the number of standard deviations from average with the segment threshold for deletions or duplications at +/-3.

Supplemental Figure S3: CNV events detected by CGH in soybean fast neutron root and nodule mutants. Full chromosome views of CNV events are depicted for all 20 soybean chromosomes. The normalized log2 ratio of sample to control data is plotted as the median across 11 probe datapoints across chromosome positions. Results from each array are color-coded for mutants RN1 through RN3. RN1 = non-nodulating. RN2 = robust, precocious pod
set. RN3 = hypernodulating. The gray overlay represents CGH data from control vs. control comparative genomic hybridizations. Colored regions above and below the control regions potentially represent copy number change differences. The y-axis scale is in terms of the number of standard deviations from average, and the segment threshold for deletions or duplications is +/-3.

**Supplemental Figure S4: Genomic regions of high variability across the soybean fast neutron population detected by CGH and SNP genotyping.** Regions of high variation in the soybean fast neutron population are displayed by chromosome and chromosome position in a 3D plot. The absolute value of the normalized log$_2$ ratio of sample to control intensities at each probe position was calculated for each of thirty CGH arrays. The median value across an eleven-probe sliding window was calculated using the average of the above at each probe position, and values that peaked above the 95$^{th}$ percentile border (1.103589) were plotted as regions of high variation detected by CGH (blue diamonds). BLAST positions of SNP markers from the Illumina Goldengate SNP genotyping platform that showed variation across the soybean fast neutron population are indicated (red bars). The majority of CGH and SNP genotyping results coincide (high stacks of blue diamonds above most red bars), indicating that these are likely regions of M92-220 intravarietal heterogeneity.

**Figure 5: Exome resequencing confirms gene deletions detected by comparative genomic hybridization.** A) The corrected log$_2$ ratios of sample PO1 to control intensities are shown for chromosome 10 where a deletion is detected at ~48.7 Mb. B) The normalized exome resequencing log$_2$ ratios of sample PO1 to control exon counts are displayed for chromosome 10. Each colored dot represents an exon in a high-confidence gene call. The color gradient indicates the lowest (red) to highest (blue) amount of read count evidence for an exon in sample PO1 compared to the control. The absence of sequence evidence for exons at ~48.7 Mb is shown and parallels the deletion found by CGH in (A). C) The corrected log$_2$ ratios of sample PO8 to control intensities are shown for chromosome 16 where a deletion is detected at ~28.1 Mb. D) The normalized exome resequencing log$_2$ ratios of sample PO8 to control exon counts are displayed for chromosome 10. Each colored dot represents an exon in a high-confidence gene call. The color gradient indicates the lowest (red) to highest (blue) amount of read count
evidence for an exon in sample PO8 compared to the control. The absence of sequence evidence for exons at ~28.1 Mb is shown and parallels the deletion found by CGH in (C).

Figure 6: Demarcation and confirmation of deletion regions and cosegregating phenotypes by PCR. A) Agarose gel electrophoresis of the PCR product across a deletion region in short trichome mutant VP1 next to 100 bp marker (M) and wild-type (WT) M92-220 template PCR control lanes. B) A diagram shows the reference sequence region length (4254 bp) versus the VP1 mutant region length (579 bp) characterized by PCR amplification and sequencing of the region. C) Agarose gel electrophoresis of the PCR product across a deletion region in high seed protein mutant PO1 next to a 1 kb marker (M) and wild-type (WT) M92-220 template PCR control lanes. D) A diagram depicts the reference sequence region length (39,806 bp) versus the PO1 mutant region length (~1 kb) characterized by PCR amplification and sequencing of the region. E) A deletion region on chromosome 17 was confirmed by PCR and mapped in VP5. This genetic marker locus in VP5 cosegregates with progeny displaying the short petiole phenotype (+) and is not found in wild-type (WT) or in progeny without the short petiole phenotype (-) shown in (F).
Table I. Summary statistics for soybean mutant viability as a function of fast neutron radiation dose. The M$_2$ generation was grown in Santiago, Chile as single seed descent from M$_1$ plants grown in St. Paul, MN, U.S.A.

<table>
<thead>
<tr>
<th>Radiation dose</th>
<th>M$_1$ plants emerged</th>
<th>M$_1$ plants producing M$_2$ seed</th>
<th>M$_2$ plants emerged</th>
<th>M$_2$ plants producing M$_3$ seed</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Gy</td>
<td>76.24%</td>
<td>48.98%</td>
<td>88.00%</td>
<td>80.00%</td>
</tr>
<tr>
<td>8 Gy</td>
<td>73.14%</td>
<td>33.80%</td>
<td>69.67%</td>
<td>57.33%</td>
</tr>
<tr>
<td>16 Gy</td>
<td>69.22%</td>
<td>23.36%</td>
<td>66.58%</td>
<td>56.83%</td>
</tr>
<tr>
<td>32 Gy</td>
<td>61.04%</td>
<td>15.48%</td>
<td>74.40%</td>
<td>54.80%</td>
</tr>
</tbody>
</table>
Table II. Seed protein and oil composition data by NIR on M6 seed from eight soybean fast neutron mutant lines (PO1-PO8) compared to the wild-type M92-220 in 2010. Mean and standard deviations are shown. Maturity rankings are on a five-point scale from earliest (5) to latest (1) dates of plant maturity.

<table>
<thead>
<tr>
<th>ID</th>
<th>% Protein ±</th>
<th>% Oil ±</th>
<th>% (Protein + Oil) ±</th>
<th>Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>M92-220</td>
<td>39.16 ± 0.52</td>
<td>21.20 ± 0.39</td>
<td>60.36 ± 0.65</td>
<td>4</td>
</tr>
<tr>
<td>PO1</td>
<td>46.99 ± 1.18</td>
<td>12.92 ± 1.19</td>
<td>59.91 ± 1.68</td>
<td>2</td>
</tr>
<tr>
<td>PO2</td>
<td>40.42 ± 0.36</td>
<td>20.59 ± 0.11</td>
<td>61.01 ± 0.38</td>
<td>3</td>
</tr>
<tr>
<td>PO3</td>
<td>49.31 ± 0.62</td>
<td>16.54 ± 0.32</td>
<td>65.84 ± 0.70</td>
<td>3</td>
</tr>
<tr>
<td>PO4</td>
<td>43.63 ± 0.35</td>
<td>10.93 ± 0.47</td>
<td>54.57 ± 0.59</td>
<td>3</td>
</tr>
<tr>
<td>PO5</td>
<td>42.37 ± 1.07</td>
<td>11.45 ± 0.54</td>
<td>53.82 ± 1.20</td>
<td>4</td>
</tr>
<tr>
<td>PO6</td>
<td>35.73 ± 0.29</td>
<td>21.68 ± 0.30</td>
<td>57.41 ± 0.42</td>
<td>1</td>
</tr>
<tr>
<td>PO7</td>
<td>42.39 ± 0.91</td>
<td>21.24 ± 0.22</td>
<td>63.63 ± 0.93</td>
<td>5</td>
</tr>
<tr>
<td>PO8</td>
<td>46.72 ± 1.42</td>
<td>14.95 ± 0.30</td>
<td>61.66 ± 1.45</td>
<td>2</td>
</tr>
</tbody>
</table>
SUPPLEMENTAL MATERIAL

*Bolon_FNSupplFigures.pdf*

Supplemental Figures S1-S4

*Bolon_FNSupplTable1.xls*

Supplemental Table S1: Mean, standard deviation and range of FN mutant M3 seed composition traits derived from three locations analyzed by NIR on a dry matter basis.

*Bolon_FNSupplTable2.xls*

Supplemental Table S2. List of deleted genes on chromosome 10 of M23.

*Bolon_FNSupplTable3.xls*

Supplemental Table S3: Universal SNPs on the Illumina Goldengate platform with evidence for segregation within the soybean fast neutron population.

*Bolon_FNSupplTable4.xls*

Supplemental Table S4: DNA copy number variation (CNV) segment borders detected in soybean fast neutron mutants by CGH.

*Bolon_FNSupplTable5.xls*

Supplemental Table S5: List of genes within detected CNV events in soybean fast neutron mutants.

*Bolon_FNSupplTable6.xls*

Supplemental Table S6. Identifiers and pedigree for soybean FN mutant lines assayed by CGH.
Figure 1: Selected soybean fast neutron mutants with visual phenotypes. A) Fast neutron mutant VP1 (top) displays short trichomes compared to wild-type soybean (bottom). B) A chimeric mutant VP3 with altered leaf pigmentation patterns is visible among a field of other soybean mutants with normal pigmentation. C) A row of petite yellow-tinted mutants (VP4). D) The short petiole and crinkled leaf mutant (VP5). E) Wild-type root and nodules are shown at three weeks post germination. F) Mutant RN1 displays abnormal shoot-root connection and does not appear to nodulate. G) Robust mutant RN2 displays elongated internodes compared to wild-type and begins pod set precociously. H) Hypernodulating mutant RN3 displays increased nodulation compared to wild-type in (E).
Figure 2: Detection of a genomic deletion region encompassing a known deleted gene in M23. A) CGH analysis identifies a deletion at the end of chromosome 10. The Y-axis represents unaveraged log₂ ratios of M23 to Bay hybridization signals. B) Confirmation of the predicted deletion by PCR. Arrows represent flanking amplification primers located 1.5 kb from the predicted 163.6 kb deletion.
Figure 3: CNV events detected by CGH in soybean fast neutron seed protein and oil mutants. CNV events are depicted for all 20 soybean chromosomes. The normalized log2 ratio of sample to control data is plotted as the median across 11 probe datapoints across chromosome positions. Results from each array were determined by fragmentation PO1 through PO8. PO1-PO3, PO6 = high seed protein. PO4 = low seed oil and high seed raffinose and sucrose. PO5 = low seed oil. PO8 = low seed protein. PO7 = high seed oil. The gray overlay represents CGH data from control vs. control comparative genome hybridizations. Colored regions above and below the control regions potentially represent CNV events. The y-axis scale is in terms of the number of standard deviations from average with the segment threshold for deletions or duplications at +/-3.
Figure 4: CNV events detected by CGH in soybean fast neutron visual phenotype mutants. Full chromosome views of CNV events are depicted for all 20 soybean chromosomes. The normalized log2 ratio of sample to control was calculated. The graph includes 11 probe data points across chromosome positions. Results from each array are color-coded for mutants VP1 through VP9. VP1 and VP2 = short trichomes. VP3 = chimeric leaf pigmentation. VP4 = petite and yellow leaf. VP5 = short petioles and twisted leaf. VP6 = copper leaf. VP7 = abnormal floral meristem development. VP8 = fused trifoliates. VP9 = thick, twisted petioles. The gray overlay represents CGH data from control vs. control comparative genome hybridizations. Colored regions above and below the control regions potentially represent copy number change differences. The y-axis scale is in terms of the number of standard deviations from average with the segment threshold for deletions or duplications at +/-3.
Figure 5: Exome resequencing confirms gene deletions detected by comparative genome hybridization. A) The corrected log₂ ratios of sample PO1 to control intensities are shown for chromosome 10 where a deletion is detected at ~48.7 Mb. B) The normalized exome resequencing log₂ ratios of sample PO1 to control exon counts are displayed for chromosome 10. Each colored dot represents an exon in a high-confidence gene call. The color gradient indicates the lowest (red) to highest (blue) amount of read count evidence for an exon in sample PO1 compared to the control. The absence of sequence evidence for exons at ~48.7 Mb is shown and parallels the deletion found by CGH in (A). C) The corrected log₂ ratios of sample PO8 to control intensities are shown for chromosome 16 where a deletion is detected at ~28.1 Mb. D) The normalized exome resequencing log₂ ratios of sample PO8 to control exon counts are displayed for chromosome 10. Each colored dot represents an exon in a high-confidence gene call. The color gradient indicates the lowest (red) to highest (blue) amount of read count evidence for an exon in sample PO8 compared to the control. The absence of sequence evidence for exons at ~28.1 Mb is shown and parallels the deletion found by CGH in (C).
Figure 6: Demarcation and confirmation of deletion regions and cosegregating phenotypes by PCR. A) Agarose gel electrophoresis of the PCR product across a deletion region in short trichome mutant VP1 next to 100 bp marker (M) and wild-type (WT) M92-220 template PCR control lanes. B) A diagram shows the reference sequence region length (4254 bp) versus the VP1 mutant region length (579 bp) characterized by PCR amplification and sequencing of the region. C) Agarose gel electrophoresis of the PCR product across a deletion region in high seed protein mutant PO1 next to a 1 kb marker (M) and wild-type (WT) M92-220 template PCR control lanes. D) Diagram depicts the reference sequence region length (39,806 bp) versus the PO1 mutant region length (~1 kb) characterized by PCR amplification and sequencing of the region. E) A deletion region was confirmed by PCR and mapped in VP5. This genetic marker locus in VP5 cosegregates with progeny displaying the short petiole phenotype (+) and is not found in wild-type (WT) or in progeny without the short petiole phenotype (-) shown in (F).