Development and utilization of a new chemically-induced soybean library with a high mutation density

Zhongfeng Li, Lingxue Jiang, Yansong Ma, Zhongyan Wei, Huilong Hong, Zhangxiong Liu, Jinhui Lei, Ying Liu, Rongxia Guan, Yong Guo, Longguo Jin, Lijuan Zhang, Yinghui Li, Yulong Ren, Wei He, Ming Liu, Nang Myint Phyu Sin Htwe, Lin Liu, Bingfu Guo, Jian Song, Bing Tan, Guifeng Liu, Maiquan Li, Xianli Zhang, Bo Liu, Xuehui Shi, Sining Han, Sunan Hua, Fulai Zhou, Lili Yu, Yanfei Li, Shuang Wang, Jun Wang, Ruzhen Chang, and Lijuan Qiu

1. National Key Facility for Gene Resources and Genetic Improvement/Key Laboratory of Crop Germplasm Utilization, Ministry of Agriculture, Institute of Crop Science, Chinese Academy of Agricultural Science, Beijing 100081, China
2. Soybean Research Institute, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China
3. College of Agriculture, Yangtze University, Jingzhou 434025, China
*Correspondence: Lijuan Qiu (qiulijuan@caas.cn)
doi: 10.1111/jipb.12505

Abstract Mutagenized populations have provided important materials for introducing variation and identifying gene function in plants. In this study, an ethyl methanesulfonate (EMS)-induced soybean (Glycine max) population, consisting of 21,600 independent M2 lines, was developed. Over 1,000 M4 (5) families, with diverse abnormal phenotypes for seed composition, seed shape, plant morphology and maturity that are stably expressed across different environments and generations were identified. Phenotypic analysis of the population led to the identification of a yellow pigmentation mutant, gyl, that displayed significantly decreased chlorophyll (Chl) content and abnormal chloroplast development. Sequence analysis showed that gyl is allelic to MinnGold, where a different single nucleotide polymorphism variation in the Mg-chelatase subunit gene (ChlIa) results in golden yellow leaves. A cleaved amplified polymorphic sequence marker was developed and may be applied to marker-assisted selection for the golden yellow phenotype in soybean breeding. We show that the newly developed soybean EMS mutant population has potential for functional genomics research and genetic improvement in soybean.

Edited by: Chun-Ming Liu, Institute of Crop Science, Chinese Academy of Agricultural Sciences and Institute of Botany, Chinese Academy of Sciences, China
Received Aug. 15, 2016; Accepted Oct. 20, 2016; Online on Oct. 24, 2016
OO: OnlineOpen, paid by authors

INTRODUCTION

Soybean (Glycine max (L.) Merr.) is one of the most important crops in the world. It is also an important source of plant-based protein for both humans and livestock. The release of the soybean reference genome (cv. Williams 82) (G. max; Schmutz et al. 2010) opened a new era in functional genomics for this important agronomic species. However, the lack of a high-throughput transformation platform, combined with the complexity of the genome structure, has limited progress in gene discovery. Mutants play an important role in identifying gene function (Zhu et al. 2005; Gabrielson et al. 2006), and have been successfully used to study gene function in both model and non-model plant species (Cui et al. 2013). Several mutation methods...
are available for introducing genomic variation by chemical, radiation and transformation-induced mutagenesis of plant genomes. Ionizing radiation (X-rays, gamma rays and fast neutrons) induces mainly nucleotide deletions of various sizes at a relatively low density (Shirley et al. 1992; Cecchini et al. 1998), which often results in loss of function mutants (Anai et al. 2008, 2012b), which are beneficial for gene discovery but are not a good choice for identifying allelic mutations in target genes such as GBSSI, now used for crop improvement (Slade et al. 2005). Insertional mutagenesis using T-DNA and transposon tagging is a powerful tool for associating genotype with phenotype, and has been successfully used to study gene function in both crops and model plants (Pan et al. 2003; Sallaud et al. 2004; Mathieu et al. 2009) when coupled with high-throughput transformation. However, the absence of a comparable transformation system, together with the requirement for tissue culture, makes it impractical to create large mutant populations using these strategies; thus, alternative tools for identifying soybean genes are needed. Chemical mutagens are a particularly promising means of mutagenesis because of their ability to enact high mutation density randomly distributed within the genome. Chemically mutagenized populations have been generated in *Arabidopsis* (McCallum et al. 2000; Greene et al. 2003), rice (Wu et al. 2005; Till et al. 2007), wheat (Slade et al. 2005; Uauy et al. 2009), maize (Till et al. 2004; Weil and Morde 2007), barley (Caldwell et al. 2004), sorghum (Xin et al. 2008) and tomato (Menda et al. 2004; Minoia et al. 2010). In recent years, considerable efforts have been made in developing genomic resources for soybean. Carroll et al. (1985) mutagenized the seeds of soybean (cv. Bragg) with ethyl methanesulfonate (EMS) and isolated 15 independent nitrate-tolerant symbiotic (nts) mutants by screening 2,500 M₁ families. Four chemically mutagenized soybean populations were developed by treatment of seeds of the soybean (cv. Williams 82 and Forrest) with EMS and N-nitroso-N-methylurea (NMU) (Cooper et al. 2008). Anai (2012a) developed a soybean mutant population consisting of more than 10,800 M₂₃ lines, generated by treatment of three soybean landraces with two different concentrations of EMS, yielding three sub-populations. More recently, a high-density mutant library consisting of 1,477 M₃ lines in soybean was constructed, and the mutation density was evaluated based on whole-genome re-sequencing analysis of 12 independent mutant lines (Tsuda et al. 2015). These mutant collections are important resources for both functional genomics research and cultivar breeding in soybean. However, compared with the number of genes (46,430 to 55,616) (Libault et al. 2010; Schmutz et al. 2010) predicted in the soybean reference genome, there are multiple copies for 75% of the predicted genes owing to the polyploid nature of soybean genome. Further, 57% of the genomic sequence occurs in repeat-rich, low-recombination heterochromatic regions surrounding the centromeres. Therefore, the mutants currently available to the soybean community are far from sufficient for modern phenomics analysis, and more efforts are required to develop useful EMS mutant populations.

Next-generation sequencing (NGS) technology, coupled with the growing number of sequenced genomes, opens up the opportunity to redesign genotyping strategies for more effective genetic mapping and single nucleotide polymorphism (SNP) discovery. Initially, to reduce costs without compromising SNP quality, several methods were developed that involved sequencing only a small fraction of the genome. These approaches include building reduced representation libraries (RRL) (Hyten et al. 2010; Varala et al. 2011; Sun et al. 2013), utilizing restriction site-associated DNA (RAD) (Baird et al. 2008) sequencing, genotyping by sequencing (GBS) (Elshire et al. 2011; Sonah et al. 2013; Liu et al. 2014), exome sequencing (Mascher et al. 2013, 2014) or reducing sequence depth (Huang et al. 2009; Xie et al. 2010). With the decreased cost of DNA sequencing technologies, whole-genome deep re-sequencing based on NGS, including sequencing pooled DNA (Schneeberger et al. 2009; Austin et al. 2011; Abe et al. 2012; Leshchiner et al. 2012; Fekih et al. 2013), resequencing different individuals or accessions (Lam et al. 2010; Xu et al. 2011; Nordström et al. 2013; Zhou et al. 2015) and transcriptome sequencing (Trick et al. 2012; Islam et al. 2016) have been used to greatly accelerate the identification of mutagen-induced or naturally occurring mutations. An increasing number of successful examples have demonstrated the feasibility of the method by identification of EMS-induced, causal mutations in *Arabidopsis* (Ashelford et al. 2011; Hartwig et al. 2012; Leshchiner et al. 2012), rice (Abe et al. 2012), soybean (Zhou et al.
Whole-genome sequencing based on NGS is being increasingly utilized for SNP discovery and gene identification in both crops and model organisms because of lowering cost and high efficiency.

The objective of our study was to build an EMS mutant population to complement the wealth of functional genomics resources currently available for soybean. Our analysis included evaluating mutants with observed phenotypes by progeny testing, estimation of mutation frequency in the chemically mutagenized population, and the use of mutants for germplasm enhancement and gene discovery.

**RESULTS**

**Soybean EMS population development**

Approximately 80,000 seeds of soybean cv. Zhongpin661 (Zp661) were mutagenized with EMS. In the first season, all M₁ seeds were planted and 21,600 M₁ plants were harvested. A single seed descent population was developed to screen for plant morphological mutants and all M₃ seeds were collected from 10,700 independent M₂ plants (Figure 1). In contrast to developmental phenotypes observed in the M₁ generation that often result from physiological damage due to the chemical mutagen, variant phenotypes occurring in the M₂ generation were due to gene mutations. This study will facilitate the identification and utilization of these new genetic resources for soybean improvement.
generation are more likely due to heritable effects, and were used for further analysis. M3 seeds were collected from independent M2 individuals to form the basis of our EMS mutant library. Seeds from each M1 plant were also harvested for phenotypic analysis of seed shape and composition. Progeny tests were successively applied to the M2 to M5 generations for all M2 mutants with variant phenotypes.

As expected, seeds exposed to EMS solution showed reduced emergence and physiological damage, including growth inhibition of the main stem and reduced production of viable seed. Our data indicated that soybean seeds treated with 50 mmol/L EMS solution displayed a low germination rate (50%), which is similar to what has been observed in rice (Till et al. 2007) and wheat (Uauy et al. 2009). Further, we also observed a dramatic reduction in plant viability in the M1 generation, coupled with reduced fertility. In the M2 generation, the germination rate increased to 49.5% of all plants. Some abnormal phenotypes resulting from physiological damage, such as the loss of primary or trifoliate leaves and two primary stems, disappeared when we reached the M3 generation.

**Screening for seed composition mutants**

The EMS soybean population was developed to serve as a resource to identify seed-composition mutations. In total, 1,887 of 21,600 M2 lines were assayed by near infrared (NIR) spectroscopy to identify lines with changes in protein or oil content in seeds, selecting those showing extensive changes (Figure 2A, C). In M2 seeds, 141 lines were selected based on our initial screen, of which 115 lines had high protein content (46.1–65.0%), four mutants had marginal (35.0–35.7%) protein levels (Figure 2A). Interestingly, among 22 mutants identified with significant differences in seed oil content, only one showed increased seed oil content (23.1%), while it was reduced (13.3–15.7%) in the others (Figure 2C). These variations were tested from M3 in 2013 through M4 in 2014. Forty-eight M4 lines, derived from 47 M1 plants, exhibited higher seed protein content (44.1–46.4%) and very few mutants had a minor combined increase in seed protein and oil content compared with the parental line, cv. Zp661 (Figure 2B). However, the selected high/low-seed oil or low-seed protein phenotypes in M3 were not observed in the 2013 and 2014 seasons (Figure 2B,D), indicating that this phenotype may be affected by growth environment.

**Screening seed trait mutants**

Among 21,600 M2 families, a total of 323 seed shape mutant lines were identified as displaying variant phenotypes for 100-seed weight, seed hilum or coat color, seed coat shape and other seed traits. This accounted for 1.5% of the M2 lines, including 201 mutants that showed a clear increase or decrease in 100-seed weight, 33 lines that displayed dehiscence of the seed coat, and 89 families showing seed hilum or seed coat color variation (Table 1). Continuous progeny testing revealed 60 mutant lines with seed shape variation that was stably inherited from the M2 to the M4 generation. Of these, 17 M4 lines showed an extremely significant decrease in 100-seed weight compared to cv. Zp661 (Student’s t-test, P < 0.01) (Figure 3J), whereas just one line displayed a significant increase in 100-seed weight (Student’s t-test, P < 0.05) (Figure 3I). In the M4 generation, 42 mutant lines with seed coat or hilum variation were screened. The lack of M4 lines with seed coat dehiscence likely resulted from physiological damage by mutagenesis or external growth conditions.

**Screening visual morphological trait mutants**

Visible plant phenotypes were recorded for the whole EMS mutant population, expanding soybean germplasm resources. Over 4,100 independent individuals from a single-seed descent M2 population were observed to display abnormal visual phenotypes. Variant phenotypes observed in the field were divided into several categories: leaf phenotype, main stem characters, plant height, branch number, sterility, pubescence density, pod coat color and maturity (Figure 3). A summary of our phenotypic analysis is presented in Table 2. Plants with abnormal visual phenotypes account for approximately 38.6% of the M2 population. Progeny testing in the summers of 2013 and 2014 confirmed that 894 M4 (M5) lines with visible, variant phenotypes were genetically stable. Of these, 360 M4 lines derived from independent M2 individuals showed abnormal leaf phenotypes, accounting for 3.4% of the total M2 plants, 293 M4 lines with growth-period variation ranked second, and 154 M4 (M5) lines with altered plant height accounted for 1.5% of the M2 population. There were also consistent genetic variations in branch number, pod color, pubescence density and pod-setting, found in six to 27 M4 (M5) lines each, but no lines showed stable
Table 1. Progeny test of seed shape variation mutants from M2 to M4 in the field

| Variation phenotypes          | No. of plants in M3 | M3 generation (No. of lines) | M4 generation (No. of lines) | Percentagea (%) |
|-------------------------------|---------------------|------------------------------|------------------------------|-----------------|
| Large seed                   | 67                  | 2                            | 1                            | 0.005           |
| Small seed                   | 134                 | 17                           | 17                           | 0.08            |
| Seed coat/hilum color        | 89                  | 42                           | 42                           | 0.20            |
| Seed coat dehiscence         | 33                  | 1                            | 0                            | 0               |
| Total                        | 323                 | 62                           | 60                           | 0.27            |

aNumber of M4 mutant lines divided by the total number (21,600) of M3 lines, × 100.

Figure 2. The distribution of seed protein and oil content in M2 and M4 lines (A–B, protein; C–D, oil)
The parental line, Zp661 was considered as the wild-type control and is indicated in each panel by red arrows. Protein content of the parent line was determined to be 42.4 ± 1.2% and 41.1 ± 1.1% in the M4 and M2 generations, respectively. Lines considered to contain high protein content were those that had greater than 45% seed protein concentration. The oil content of the wild-type lines had an average oil content of 19.6 ± 0.6% and 20.7 ± 0.9% in the M4 and M2 generations, respectively. The cut-off threshold for lines to be considered to contain high levels of seed oil was determined to be >21%.
lodging resistance across generations. The study identified 75 M₄ (M₅) mutant lines that are potentially heterozygous, accounting for 8.4% of the total M₄ (M₅) variation families. Among them, visible variation was mainly confined to fertility and leaf phenotypes, which were readily affected by environmental factors. One or more phenotypes were recorded for 275 mutant lines. In addition, a separate screen for salt-tolerance was also conducted under salt stress conditions using a duplicate M₂ population; however, no visible mutants were identified. Root phenotypes were not recorded in the M₂ population. Detailed information regarding our phenotypic analysis can be found online at http://www.cgris.net/query/croplist.php# (soybean).

**Mutation density in the EMS population**

To estimate the mutation density in our population, the genomes of four randomly selected mutant lines with different phenotypes (yellowish plant, dwarf, male-sterile mutant and the wild-type parent line cv. Zp661) were re-sequenced using 101-base paired-end sequencing (Illumina HiSeq 2000 sequencer), yielding 22.07–43.97 Gb for these samples (Table 3). Over 97% of the reads were uniquely mapped to the Williams 82 reference genome,
corresponding to \( >6 \times \) coverage of the genome (980 Mb). Based on alignment to the Williams 82 reference genome, 70,032–93,530 SNPs were identified between the mutant lines and Zp661, corresponding to an estimated mutation density of \( 1/11.8 \) kb on average. Among the four resequenced mutant lines, 11,899–58,628 SNPs were detected between each other and 8,015 SNPs were shared in different genomes (Figure 4), indicating that a large number of SNP variations are present in the mutated genomes.

### Identification of a chlorophyll-deficient soybean EMS mutant

Mutants that displayed abnormalities in pigmentation accounted for 0.25% of the \( M_2 \) generation. Notably, one mutant (\( \text{gyl} \)) had golden yellow trifoliate leaves, pod coat and cotyledons in addition to reduced plant height, relative to the wild type throughout all developmental stages (Student’s t-test, \( P < 0.01 \)). Pigment assays showed that the total chlorophyll (Chl) content of \( \text{gyl} \) was consistently only 60% that of the wild-type parent cv. Zp661, with 50% and 90% reductions of Chl \( a \) and Chl \( b \) content compared to those of the wild type, respectively. Interestingly, \( \text{gyl} \) showed a substantial increase in the Chl \( a/b \) ratio (3.2-fold), likely due to sharper declines in Chl \( b \) relative to Chl \( a \) synthesis. To determine whether the \( \text{gyl} \) mutation affected chloroplast development, the ultrastructure of chloroplasts of both the \( \text{gyl} \) mutant and wild-type plants at seedling stage was observed by transmission electron microscopy. Abnormal non-appressed thylakoid stacking was observed in \( \text{gyl} \), with only a few rare, granal stacks observed (Figure 5A), whereas the wild type had several large granal stacks (Figure 5B).

Genetic mapping of the \( \text{GYL} \) gene was performed using the \( F_2 \) population from a cross of Jidou12 × \( \text{gyl} \). Using bulked segregant analysis (BSA), we mapped the

| Group       | Subgroup                  | No. of mutants | Ratio of the mutants (%) | Heterozygous mutants (\( M_{(45)} \)) | Percentage (%) |
|-------------|---------------------------|----------------|--------------------------|--------------------------------------|----------------|
| Leaf        | Number of leaflets        | 31 22 14       | 0.13                     | 4 11                                   | 28.6           |
|             | Yellow leaf               | 56 44 26       | 0.25                     | 0 3                                    | 0.0            |
|             | Leaf size variation       | 587 228 144    | 1.36                     | 11 10                                  | 7.6            |
|             | Leaflet shape             | 115 82 63      | 0.59                     | 3 4                                    | 4.8            |
|             | Curled leaf               | 270 132 112    | 1.06                     | 10 8                                   | 8.9            |
|             | Short petiole             | 1 1 1          | 0.01                     | 0 0                                    | 0.0            |
| Main stems  | Strong stem               | 48 5 0         | 0                        | 0 0                                    | 0.0            |
|             | Thin stem                 | 193 6 6        | 0.06                     | 4 66.7                                 | 11.8           |
|             | Double stems              | 20 0 0         | 0                        | 0 0                                    | 0.0            |
|             | Loss of main stem         | 33 0 0         | 0                        | 0 0                                    | 0.0            |
|             | Twisted stems             | 28 9\( (M_4) \) 9\( (M_5) \) | 0.08 | 9 100.0 | 0.0 |
|             | Growth habit              | 29 2\( (M_4) \) 2\( (M_5) \) | 0.02 | 1 50.0 | 0.0 |
|             | Others                    | 9 0 0         | 0                        | 0 0                                    | 0.0            |
| Plant height| Dwarf plant               | 353 156 140    | 1.32                     | 0 0                                    | 0.0            |
|             | Tall plant                | 66 14\( (M_4) \) 14\( (M_5) \) | 0.13 | 7 50.0 | 0.0 |
| No. of branches| Increased branches       | 184 30 8      | 0.08                     | 7 87.5                                 | 0.0            |
| Pod         | Pod color/shape           | 23 18 10       | 0.09                     | 1 10.0                                 | 0.0            |
| Pubescence  | Length/color/ density     | 50 39 27       | 0.26                     | 3 11.1                                 | 0.0            |
| Sterility   | Partial sterility         | 125 38 25      | 0.24                     | 15 60.0                                | 0.0            |
|             | Sterility                 | 266 0 0        | 0                        | 0 0                                    | 0.0            |
| Lodging     | Prone to lodging          | 55 0 0         | 0                        | 0 0                                    | 0.0            |
| Growth period| Early maturity            | 375 136 39     | 0.37                     | 0 100.0                                | 0.0            |
|             | Late maturity             | 1,223 575 254  | 2.40                     | 0 100.0                                | 0.0            |
| Total       |                           | 4,140 1,537 894 | 8.44                    | 75 8.4                                 | 0.0            |

\( ^a \)Percentage (%) was calculated as \( 100 \times \) number of final mutant lines for each phenotype divided by 10,700 \( M_2 \) plants. \( ^b \)Percentage (%) was calculated as \( 100 \times \) number of final heterozygous mutant lines divided by the final mutant lines for each phenotype.
GYL gene to a 1 Mb region between two simple sequence repeat (SSR) markers (BARCSOYSSR_13_1412 and BARCSOYSSR_13_1463) on chromosome 13. Coincidently, a known chlorophyll (Chl) synthesis-associated gene, Glyma.13g232500 (the Glyma.Wm82.a2.v1 annotation) was located within the same region. Glyma.13g232500 encodes Mg-chelatase subunit ChlI1a, identified previously in the chlorophyll-deficient mutant, MinnGold, which has been shown to have yellowing of aerial tissues (Campbell et al. 2014). Four pairs of overlapping primers spanning the genomic region of Glyma.13g232500 were designed to assay the gyl genotype. Sequence analysis identified a nonsynonymous SNP mutation (G589A) in the third exon of Glyma.13g232500, leading to a missense mutation resulting in glutamic acid with a lysine residue (E268K) in the deduced amino acid sequence. This finding indicates that gyl is allelic to MinnGold, which results from a nonsynonymous SNP (G605A) mutation in the third exon. The linkage relationship between the SNP genotype and the mutant phenotype was confirmed by genotyping F2 individuals derived from a cross between gyl and Zp661. Of 169 F2 progeny, 85 F2 individuals with the wild-type phenotype showed homozygous G/G or heterozygous mutations G/A at the SNP locus (G589A), whereas all 84 F2 individuals with the mutant-type phenotype displayed a

Figure 4. Overview of single nucleotide polymorphism (SNP) data generated analyzing the four resequenced mutant lines

Venn diagram comparison of SNPs identified in four phenotypically distinct M4 (or M5) mutants: Ems-1 (yellow leaves), ems-2 (male-sterile), ems-3 (dwarf) and ems-4 (visibly normal).
homozygous mutant-type SNP (A/A). Next, we designed a cleaved amplified polymorphic sequence (CAPS) marker based on the identified gyl SNP locus (G589A), using the restriction enzyme Mbo II (GAAGA (8/7)), to evaluate its utility for marker-assisted selection in soybean breeding. In addition to the gyl mutant, five wild-type soybean accessions including Peking, Jidou12, Zhonghuang13, Zhonghuang39 and ZYD3687, were analyzed by the CAPS marker. The targeted fragment from all five wild-type lines could be detected as two fragments of 254 and 66 bp when Mbo II cleaved polymerase chain reaction (PCR) amplification products were separated by 2.0% agarose gel electrophoresis (Figure 6). In contrast, the amplified product from the gyl mutant does not contain an Mbo II restriction, and is detected as a single band of 320 bp. This result supports our data that demonstrate that the G589A SNP in Glyma.13g232500 is responsible for the abnormal pigmentation phenotype in gyl and the developed CAPS can be utilized in soybean breeding.

**DISCUSSION**

Chemically mutagenized populations of soybean (Cooper et al. 2008) as well as wheat (Slade et al. 2005; Uauy et al. 2009) show higher mutation frequencies than that of barley (Caldwell et al. 2004), rice (Till et al. 2007), maize (Till et al. 2004) and Arabidopsis (Greene et al. 2003). In addition, the mutation frequency can be increased without adverse effects, owing to the genetic redundancy provided by duplicated genes and/or the polyploidy of the plant genome.

In this study, we developed a soybean EMS-treated population to screen various phenotypic mutants. The estimated mutation density in our new soybean mutant population is approximately five times higher than that reported by Tsuda et al. (2015). In our study, estimated mutation density (~1/11.8 kb, 50 mmol/L EMS solution) is higher than those reported for other EMS mutant populations in soybean (~1/140~1/550 kb, Cooper et al. 2008; 0.3~1.3/Mb, Anai 2012a). The high observed mutation density will provide more allelic mutations.
of target genes, in addition to knockout mutants that are valuable for reverse genetics analysis. According to prior studies by the Seattle TILLING Project (Till et al. 2003), our estimated mutation density of ~1/11.8 kb is suitable to develop efficient screening platforms to identify allelic mutant variants underlying important agronomic traits with TILLING.

Mutagenized populations are valuable sources of genetic and phenotypic diversity. In this study, we have described the identification and characterization of several M4/M5 lines with a range of diverse phenotypes, including seed composition, seed shape and plant morphology that are stably inherited across environments and generations, which has been problematic in other mutagenized populations used for reverse genetics. Moreover, some lines display agronomically important traits such as early maturity, dwarf and yellow-leaf mutants, and can be used as germplasm resources in soybean breeding. Since soybean is a key crop of modern agriculture owing to its high protein and oil content of its seed, some mutants that we have identified can potentially be used directly for development of high-quality commercial soybean varieties. Screening for cold tolerance (Schor et al. 1993), soybean cyst nematode resistance (Cook et al. 2012) palmitic acid content in seeds (Pham et al. 2010; Anai et al. 2012b), and nodulation (Carroll et al. 1985) will be carried out in future experiments.

Of the identified mutant lines in our study (http://www.cgris.net/query/croplist.php# (soybean)), most are homozygous, which will accelerate genetic approaches to isolate genes of interest underlying important agronomic traits. However, only a paucity of genes related to Chl biosynthesis have been definitively identified from soybean (Campbell et al. 2014). In our study, 26 M4 lines with reduced pigment levels were isolated. Molecular characterization of the underlying genes responsible for the pigmentation mutants identified in our mutant population will yield new insight into our understanding of chlorophyll biosynthesis in soybean. Further, the golden yellow phenotype has the potential to be used as a morphological marker for producing hybrid seeds by elimination of nonrecombinant individuals in seed production (Wu et al. 2002). Our results provide a new community resource for reverse genetics, marker development and molecular breeding in soybean.

**MATERIALS AND METHODS**

**Mutagenesis, population development and mutant screening**

Eighty thousand seeds derived from the soybean (Glycine max) cv. Zhongpin661 (Zp661), widely used in breeding, were imbibed in double-distilled water containing 50 mmol/L EMS phosphate solution for 9 h at room temperature with gentle shaking. Seeds were rinsed with running tap water for 2 h then air-dried overnight in nylon net bags. M1 seeds were planted in the field in Shunyi, Beijing, and M2 seeds were harvested from more than 21,600 individual M1 plants (Figure 1). The M3 population was propagated by single-seed descent to screen for mutants with abnormal visible phenotypes in the winter of 2012 in Hainan province, China. M3 seeds collected from each M1 individual were analyzed for seed composition by NIR spectroscopy (Bruker MPA, Germany), observed for variation in seed shape, and then the lines with abnormal seed-composition mutations were planted in single rows in the summer of 2013 in Beijing, China. M3 seeds were harvested separately from each of more than 53,000 M2 plants. Twenty to 40 M3 progeny were planted in the field for each individual M2 line. Plants were scored for visible phenotypes every 2 weeks following seedling emergence until maturity. All mutants with abnormal visual phenotypes, compared to their parental line cv. Zhongpin661, were cataloged on our searchable web-based database http://www.cgris.net/query/croplist.php# (soybean). Until now, the identified mutant lines have been given the unified numbers from ZDD25362 to ZDD26133.

**Segregation population, gene mapping and cultivars used for gene identification**

For the identification of a chlorophyll-deficient EMS mutant, an F2 segregating population, derived from a cross between the gyl mutant (ZDD25362) and soybean cultivar Jidou12, was grown in the greenhouse (Institute of Crop Science, Chinese Academy of Agricultural Science, Beijing, China). The parental lines and 10 random F2 mutant plants were selected and used for bulked segregant analysis. Five hundred and forty-three SSR markers (http://soybase.org/dlpages/#ssrprimer) were selected for comparative analysis of the gyl mutant and JD12. To analyze the correlation between the SNP G589A and the mutant phenotype, an
additional F₂ population, derived from a cross between the gyl (yellow leaf and cotyledons) and the wild parent Zhongpin661 (Zp661, green leaf and cotyledons) was also constructed. In addition, one wild (ZYD3687) and three cultivated soybean accessions (Peking, Zhonghuang13 and Zhonghuang39) were also used to evaluate the CAPS markers.

**Pigment determination and transmission electron microscopy (TEM)**

Leaf samples from wild-type and gyl plants were collected from 2-week-old (14 d after germination (DAG)) plants grown in a growth chamber at medium light intensity (18 h daylight/6 h darkness). Leaf sections were initially fixed in a solution of glutaraldehyde (2.5% in 100 mmol/L cacodylate buffer, pH 7.4), followed by OsO₄ (1% v/v in 100 mmol/L cacodylate buffer, pH 7.4). Tissues were stained with uranyl acetate, dehydrated in ethanol and embedded in Spurr’s medium prior to thin sectioning. Samples were stained again and examined with a Hitachi H-7650 (Japan) TEM (Tanaka et al. 2003). Chlorophyll (Chl) and carotenoid (Cars) content were measured according to the methods by Arnon (1949), with minor modifications. Briefly, equal weights of freshly collected middle leaves from the young seedlings (14 DAG) were incubated in ethanol (95%) for 48 h in the dark. Residual plant debris was removed by centrifugation, and the supernatants were analyzed with a DU 800 UV/Vis spectrophotometer (Beckman Coulter, Brea, CA, USA) at 665, 649 and 470 nm, respectively.

**DNA extraction, primer design and PCR**

DNA was extracted from fresh leaves using the CTAB method (Saghai-Maroof et al. 1984) with minor modifications. Primers were designed using the Primer3 Input at http://primer3.ut.ee/ based on the Williams 82 genomic sequence. PCR reactions (25 μL) consisted of genomic DNA (100 ng), PCR buffer (1X), deoxynucleotide triphosphate (2 mmol/L), MgSO₄ (25 mmol/L), forward and reverse primers (2 μmol/L), Kod-Plus-Neo (0.5 U) DNA polymerase (TOYOBO, Japan) and sterile water. PCR reaction conditions involved denaturation at 94°C for 2 min, 38 cycles of denaturation at 98°C for 15 s, annealing at 58°C for 20 s, extension at 68°C for 50 s and a final extension at 68°C for 10 min before cooling to 10°C. PCR products were separated by agarose gel electrophoresis and products were visualized in a UV light box after staining with ethidium bromide.

**Genotyping and development of CAPS marker**

Four pairs of overlapping primers, including SNP133-1, SNP133-2, SNP133-3 and SNP133-4, covering the whole genomic sequence of Glyma.13g232500 were developed to assay the gyl mutant allele. PCR products were generated, sequenced and analyzed by Basic Local Alignment Search Tool analysis at http://multalin.toulouse.inra.fr/multalin/ (Corpet 1988). To develop CAPS markers, an additional primer set, SNP13g-3, was designed for a second round of amplification, using PCR products from the SNP133-4 primer set as a template. The resultant DNA fragments were digested with the restriction enzyme Mbo II (GAAGA (8/7)). The following primer sequences were used for PCR and sequencing of the population and putative mutants: SNP133-4_5’-TGTTGTTGTTTGAGTTGTT-3’ and that of the reverse primer SNP133-4 was 5’-TGGACATCCACC-GAAGTTT-3’. For the primer set SNP13g-3, the left strand was 5’-TAGAGTGTGGAAACGGATT GAC-3’ and the right strand was 5’-GCTCAGCATCCTAAACAGT-3’. PCR products were separated by 2% agarose gel electrophoresis and the target band was recovered and purified and then sequenced.

**Whole-genome resequencing and SNP detection**

Four mutant lines, derived from the EMS mutagenized population of cv. Zhongpin661, were selected for whole genome sequencing: A yellow leaf mutant (M₄, ZDD25362) with a dramatic reduction in total chlorophyll (Chl) content, a dwarf mutant (M₄, ZDD25365), a male-sterile mutant (M₅, ZDD25366) and a M₄ individual that appears to be phenotypically wild type (no uniform number). DNA samples were extracted from leaves of wild type (Zp661) and each of the four mutant lines (Abe et al. 2012). Libraries for sequencing were prepared from 5 μg DNA samples. The libraries were sequenced on the Illumina HiSeq 2000 sequencer following the manufacturer’s instructions (Zhou et al. 2015). Raw reads were filtered to eliminate sequencing errors. Adaptor sequences, reads with low-quality bases (N for > 10%), those with 50% or more bases having Phred-scaled quality score (Q-score) lower than or equal to 10, and homopolymers were trimmed/filtered from the raw data. Further, all reads were eliminated with a PHRED quality (Q) score < 20. After
data pre-processing, clean reads were aligned to the Williams 82 reference sequence using BWA (Li and Durbin 2009) software, and the aligned short reads were filtered with Coval to improve SNP calling accuracy. SNP identification was performed using the Genome Analysis Toolkit (GATK, McKenna et al. 2010) and SAMtools (Li et al. 2009). A detailed description of the protocol we used is present at the GATK website (https://www.broadinstitute.org/gatk/guide/best-practices?bpm=DNAseq#variant-discovery-ovw).

ACKNOWLEDGEMENTS

This work was supported by National Key R&D Program for Crop Breeding (2016YFD0100201), the Crop Germplasm Resources Protection (2014NWB030, 2015NWB030-05), Platform of National Crop Germplasm Resources of China (2014-004, 2015-004) and The Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS).

AUTHOR CONTRIBUTIONS

Lijuan Qiu designed the experiment, supervised the study and revised the manuscript. Zhongfeng Li performed research and wrote the manuscript. Lingxue Jiang and Zhongfeng Li mutagenized the seeds of cv. Zp661. Lingxue Jiang and Yansong Ma organized the harvest of all M1 plants. Zhongfeng Li, Jinhui Lei and Ying Liu performed phenotypic analysis of the M2 population in the winter of 2012. Zhongfeng Li and Zhongyan Wei carried out seed composition analysis. Zhongfeng Li screened seed trait mutants. Huiulong Hong and Zhangxiong Liu managed field research management and plant propagation. Zhongfeng Li and Jun Wang performed data analysis. The following researchers carried out the planting and harvesting of the EMS mutant population: Ruzhen Chang, Huiulong Hong, Zhangxiong Liu, Rongxia Guan, Yong Guo, Longguo Jin, Lijuan Zhang, Yinghui Li, Yulong Ren, Wei He, Ming Liu, Nang Myint Phyu Sin Htwe, Lin Liu, Bingfu Guo, Jian Song, Bing Tan, Guifeng Liu, Maiquan Li, Xianli Zhang, Bo Liu, Xuehui Shi, Sining Han, Sunan Hua, Fulai Zhou, Lili Yu, Yanfei Li and Shuang Wang. All authors read and approved the final manuscript.

REFERENCES

Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S, Terauchi R (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 30: 174–178

Anai T (2012a) Potential of a mutant-based reverse genetic approach for functional genomics and molecular breeding in soybean. Breed Sci 61: 462–467

Anai T, Hoshino T, Imai N, Takagi Y (2012b) Molecular characterization of two high-palmitic-acid mutant loci induced by X-ray irradiation in soybean. Breed Sci 61: 631–638

Anai T, Yamada R, Hideshima T, Kinoshita SM, Rahman, Takagi Y (2008) Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by GmFAD2-1A. Breed Sci 58: 447–452

Arnon DI (1949) Copper enzymes in isolated chloroplasts: Polyphenoloxidase in beta vulgaris. Plant Physiol 24: 1–15

Ashelford K, Eriksson ME, Allen CM, D-Amore R, Johansson M, Gould P, Kay S, Millar AJ, Hall N, Hall A (2011) Full genome resequencing reveals a novel circadian clock mutation in Arabidopsis. Genome Biol 12: 1–12

Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, McCourt P, Guttmann DS (2011) Next- generation mapping of Arabidopsis genes. Plant J 67: 715–725

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3: e3376

Caldwell DG, Mccallum N, Shaw P, Muehlbauer GJ, Marshall DF, Waugh R (2004) A structured mutant population for forward and reverse genetics in barley (Hordeum vulgare L.). Plant J 40: 143–150

Campbell BW, Mani D, Curtin SJ, Slattery RA, Michno JM, Ort DR, Schaus PJ, Palmer RG, Orf JH, Stupar RM (2014) Identical substitutions in magnesium chelatase paralogs result in chlorophyll- deficient soybean mutants. G3 (Bethesda, Md.) 5: 123–131

Carroll BJ, Mcneil DL, Gresshoff PM (1985) Isolation and properties of soybean (Glycine max (L.) Merr.) mutants that nodulate in the presence of high nitrate concentrations. Proc Natl Acad Sci USA 82: 4162–4166

Cecchini E, Mulligan BJ, Covey SN, Milner JJ (1998) Characterization of gamma irradiation-induced deletion mutations at a selectable locus in Arabidopsis. Mutat Res 401: 199–206

Cook DE, Geon LT, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, Diers BW, Jiang J, Hudson ME, Bent AF (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science 338: 1206–1209
Cooper JL, Till BJ, Laport RG, Darlow MC, Kleffner JM, Jamai A, El-Mellouki T, Liu S, Ritchie R, Nielsen N, Bilyeu KD, Meksem K, Comai L, Henikoff S (2008) TILLING to detect induced mutations in soybean. *BMC Plant Biol* 8: 1–10

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucl Acids Res* 16: 10881–10890

Cui Y, Barampuram S, Stacey MG, Hancock CN, Findley S, Mathieu M, Zhang Z, Parrott WA, Stacey G (2015) Tnt1 retrotransposon mutagenesis: A tool for soybean functional genomics. *Plant Physiol* 161: 36–47

Elshtor RJ, Glauzbite JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379

Fekih R, Takagi H, Tamiru M, Abe A, Natsume S, Yaegashi H, Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379

Gabrielson KM, Cancel JD, Anderson H, King MJ, Stottmann RW, Garnaas and selection. *Theor Appl Genet* 129: 1071–1086

Islam MS, Zeng L, Thyssen GN, Delhom CD, Kim HJ, Li P, Fang DD (2016) Mapping by sequencing in cotton (*Gossypium hirsutum*) line MD52ne identified candidate genes for fiber strength and its related quality attributes. *Theor Appl Genet* 129: 1071–1086

Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Li J, Jian M, Wang J, Shao G, Wang J, Sun SM, Zhang C (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet* 42: 1053–1059

Leshchiner I, Alexa K, Kelsey P, Adzhubei I, Austin-Tse CA, Cooney JD, Anderson H, King MJ, Stottmann RW, Garnaas MK, Ha S, Drummond IA, Paw BH, North TE, Beier DR, Goessling W, Sunyaev SR (2012) Mutation mapping and identification by whole-genome sequencing. *Genome Res* 22: 1541–1548

Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25: 1754–1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J (2009) The sequence alignment-map format and samtools. *Bioinformatics* 25: 2078–2079

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, 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

Liu H, Mayer B, Druka A, Russell JR, Hackett CA, Poland J, Ramsay L, Hedley PE, Waugh R (2014) An evaluation of genotyping by sequencing (GBS) to map the *Breviarista-tum-e* (ari-e) locus in cultivated barley. *BMC Genomics* 15: 1–11

Mascher M, Jost M, Kuo JE, Himmelbach A, Albagal A, Beier S, Scholz U, Craner A, Stein N (2014) Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol* 15: 29–42

Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Aylng S, Stueermagel B, Pfeifer M, D’Ascenzo M, Akhunov ED, Hedley PE, Gonzales AM, Morrell PK, Kilian B, Blattner FR, Scholz U, Mayer KF, Flavell AJ, Muelbairauer GJ, Waugh R, Jeddlelo JA, Stein N (2013) Barley whole exome capture: A tool for genomic research in the genus Hordeum and beyond. *Plant J* 76: 494–505

Mathieu M, Winters EK, Kong F, Wan J, Wang S, Eckert H, Luth D, Paz M, Donovan C, Zhang Z, Somers D, Wang K, Nguyen H, Shoemaker RC, Stacey G, Clemente T (2009) Establishment of a soybean transposon-based mutagenesis repository. *Planta* 229: 279–289

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol* 123: 439–442

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–1303

Menda N, Muttal P, Yeled P, Eshed Y, Zamir D (2004) In silico screening of a saturated mutation library of tomato. *Plant J* 39: 861–872

Minoia S, Petrozza A, D’Onofrio O, Piron F, Mosca G, Sozio G, Cellini G, Bendahmane A, Carriero F (2010) A new mutant genetic resource for tomato crop improvement by TILLING technology. *BMC Res Notes* 3: 69

Nordström KJ, Albani MC, James GV, Gutjahr C, Hwang B, Turck F, Paszkowski U, Coupland G, Schneeberger K (2013) Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. *Nat Biotechnol* 31: 325–330
Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen HA, Schor A, Fossati A, Soldat A, Stamp P (1993) Cold tolerance in wheat crop improvement by TILLING. BMC Plant Biol 7: 19

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4: 12

Trick M, Adamski NM, Mugford SG, Jiang C, Febrer M, Uauy C (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biol 12: 14

Varala K, Swaminathan K, Li Y, Hudson ME (2011) Rapid genotyping of soybean cultivars using high throughput sequencing. PLoS ONE 6: e24811

Well CF, Monde RA (2007) Getting the point-mutations in maize. Crop Sci 47: S60–S67

Wu D, Shu Q, Xia Y (2002) In vitro mutagenesis induced novel thermo/photoperiod-sensitive genic male sterile indica rice with green-revertible xanthan leaf color marker. Euphytica 123: 195–202

Wu J, Wu C, Lei C, Baraoidan M, Ramesh-Pamplona M, Mauleon R, Portugal A, Ulat VJ, Bruskiewich R, Wang G, Leach J, Khush G, Leung H (2005) Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant Mol Biol 59: 85–97

Xie W, Feng Q, Yu H, Huang X, Zhao Q, Xing Y, Yu S, Han B, Zhang Q (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci USA 107: 10578–10583

Xin Z, Wang M, Barkley NA, Burow G, Franks C, Pederson G, Burke J (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8: 103

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2011)
Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol* 30: 105–111

Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, Fang C, Shen Y, Liu T, Li C, Li Q, Wu M, Wang M, Wu Y, Dong Y, Wan W, Wang X, Ding Z, Gao Y, Xiang H, Zhu B, Lee SH, Wang W, Tian Z (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* 33: 408–414

Zhu Y, Hui F, Guo D, Xiao F, Shu C, Kuai B (2005) Characterization of a novel developmentally retarded mutant (*dmr1*) associated with the autonomous flowering pathway in *Arabidopsis*. *Cell Res* 15: 133–140