2017

Soybean Functional Genomics: Bridging the Genotype-to-Phenotype Gap

Jamie A. O'Rourke
United States Department of Agriculture, jorourke@iastate.edu

Michelle A. Graham
Iowa State University, magraham@iastate.edu

Steven A. Whitham
Iowa State University, swhitham@iastate.edu

Soybean Functional Genomics: Bridging the Genotype-to-Phenotype Gap

Abstract
Technological advances coupled with the economic importance of soybean have led to increased efforts to understand gene function and associate genes with phenotypes of agronomic and fundamental interest. Functional genomics approaches aim to develop sufficient understanding needed to bridge the genotype-to-phenotype gap. In general terms, functional genomics approaches begin by using highly parallelized methods to analyze genomes, transcriptomes, proteomes, and metabolomes to generate hypotheses about genes that control phenotypes. Candidate genes are then tested for their contributions to phenotypes through various methods such as RNA silencing, genetic mutation, or overexpression. In this chapter, we review the current approaches, tools, and resources that are being applied for functional genomics research in soybean.

Disciplines
Agricultural Economics | Agricultural Science | Agronomy and Crop Sciences | Plant Breeding and Genetics | Plant Pathology

Comments
This is a chapter from O'Rourke, Jamie A., Michelle A. Graham, and Steven A. Whitham. "Soybean Functional Genomics: Bridging the Genotype-to-Phenotype Gap." In The Soybean Genome, pp. 151-170. Springer, Cham, 2017. doi: 10.1007/978-3-319-64198-0_10.

Soybean Functional Genomics: Bridging the Genotype-to-Phenotype Gap

10

Jamie A. O'Rourke, Michelle A. Graham and Steven A. Whitham

Abstract

Technological advances coupled with the economic importance of soybean have led to increased efforts to understand gene function and associate genes with phenotypes of agronomic and fundamental interest. Functional genomics approaches aim to develop sufficient understanding needed to bridge the genotype-to-phenotype gap. In general terms, functional genomics approaches begin by using highly parallelized methods to analyze genomes, transcriptomes, proteomes, and metabolomes to generate hypotheses about genes that control phenotypes. Candidate genes are then tested for their contributions to phenotypes through various methods such as RNA silencing, genetic mutation, or overexpression. In this chapter, we review the current approaches, tools, and resources that are being applied for functional genomics research in soybean.

10.1 Introduction

The current assembly of the Williams 82 soybean reference genome (Wm82.a2.v1) contains 56,044 protein-coding loci and 88,647 mRNA transcripts (Schmutz et al. 2010); http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax). In addition, hundreds of loci encoding regulatory non-coding RNAs, including microRNAs (miRNAs) and phased small interfering RNAs (phasiRNAs), are present in the soybean genome

J.A. O'Rourke · M.A. Graham
United States Department of Agriculture—Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA 50011, USA

J.A. O'Rourke · M.A. Graham
Department of Agronomy, Iowa State University, Ames, IA 50011, USA

S.A. Whitham (✉)
Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA 50011, USA
e-mail: swhitham@iastate.edu

H.T. Nguyen and M.K. Bhattacharyya (eds.), *The Soybean Genome*, Compendium of Plant Genomes, DOI 10.1007/978-3-319-64198-0_10

(Arikit et al. 2014). Genome sequences of hundreds of accessions of *Glycine max* and allied species are available, which is expected to identify potentially useful genes that are not present in the reference genome (Kim et al. 2010a, b; Lam et al. 2010; Li et al. 2013, 2014b; Chung et al. 2014; Zhou et al. 2015). A major challenge facing soybean researchers is determining the functions of protein-coding and non-coding genes and understanding their contributions to phenotypes that are of agronomic importance and fundamental interest.

Functional genomics is an approach that seeks to bridge the gap between genotype and phenotype by assigning functions to genes based on a variety of experimental evidence that builds upon the availability of the genome sequence. Functional genomics encompasses many types of experiments and datasets that include highly parallelized analyses of gene expression at the mRNA (transcriptome) and protein (proteome) levels, or accumulation of small RNA species [microRNAs (miRNAs) and short-interfering RNAs (siRNAs)], and metabolites (metabolome). Such analyses seek to correlate genes with traits/phenotypes. However, they typically cannot establish causality, which requires altering the function of the gene(s) of interest. Therefore, perturbing the function of candidate genes is another main component of functional genomics studies.

In this chapter, we highlight recent studies and current approaches for soybean functional genomics encompassing genomics, transcriptomics, and gene knockout and knockdown strategies. We focus on recent developments related to tools, resources, and approaches, and we discuss how these may be used to investigate genes underlying phenotypes of interest. For recent in-depth reviews on functional genomics applied to specific topics such as root hair cells and interactions with the environment (abiotic and biotic factors), we refer readers to other reviews (Tran and Mochida 2010; Hossain et al. 2015; Liu et al. 2015; Whitham et al. 2016b).

10.2 Genome-Based Resources for Soybean Improvement

The completion of the *G. max* soybean genome in 2009 (Schmutz et al. 2010) helped facilitate the integration of vast amounts of genetic, phenotypic, and genomics data. Websites such as SoyBase (Cannon et al. 2012), SoyKB (Joshi et al. 2012), and Phytozome (Goodstein et al. 2011) allow users to access data by sequence, gene names, markers and traits of interest, expression patterns, or homology to known genes in other species. For a full description of these databases, we refer readers to the corresponding publications and websites.

Initially, the reference genome was used to target relatively simple traits controlled by single dominant genes or major quantitative trait loci (QTL). Gene silencing or mutagenesis would yield easily measurable phenotypic differences. For example, functional genomics approaches were used to target candidate genes for resistance (Meyer et al. 2009; Pandey et al. 2011; Liu et al. 2012; Cook et al. 2014), abiotic stress tolerance (Atwood et al. 2014), silencing (Curtin et al. 2011), or chlorophyll content (Zhang et al. 2009). The reference genome was also used to develop a comparative genome hybridization (CGH) array for soybean that could be used to identify structural variants within (Haun et al. 2011) and between soybean accessions (McHale et al. 2012; Anderson et al. 2014). Interestingly, structural variants between accessions of soybean were often associated with biotic stress response genes.

The soybean reference genome, combined with the cost-effectiveness of high-throughput sequencing, accelerated the use of genome-wide mRNA transcriptome sequencing (RNA-Seq) analyses. Transcriptome studies provide powerful datasets for functional genomics studies, particularly in combination with whole genome re-sequencing and methylome sequencing. Depending on the type of sequencing platform used, RNA-Seq data can provide information

ranging from the basic gene sequence, alternative splice variants, gene expression profiles, and polymorphic sites that can be used to develop simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers. RNA-Seq technology can facilitate both forward and reverse genetic screens, as candidate genes underlying a phenotype of interest can be uncovered and/or the impact of a specific sequence mutation on gene expression profiles can be examined.

Transcripts can be mapped to an existing reference genome or used to develop a de novo transcriptome assembly. Though there are a wide variety of alignment programs available, the TopHat (Trapnell et al. 2009) program has proven particularly useful in aligning RNA-Seq reads to complex genomes while accounting for splice junctions. These programs count the number of reads from each sample or replicate that align to each predicted gene. Counts can then be utilized by programs including Cufflinks (Trapnell et al. 2012), EdgeR (Robinson et al. 2010; McCarthy et al. 2012), and DESeq (Anders and Huber 2010) to identify suites of differentially expressed genes between samples. Severin et al. (2010) and Libault et al. (2010) developed gene atlases providing soybean gene expression measurements in different tissues. The Severin et al. (2010) atlas focuses largely on aboveground plant tissues including leaves, flowers, pods, and seven stages of seed development as well as roots and nodules. The Libault et al. (2010) atlas focuses largely on belowground tissues including root hair cells, root tips, roots, mature nodules, and it includes aboveground tissues—leaves, shoot apical meristems, flowers, and green pods. Arikit et al. (2014) generated an atlas of small RNAs developed from 69 libraries representing vegetative and reproductive tissues. These studies paved the way for soybean researchers studying gene expression changes in response to abiotic stress (Peiffer et al. 2012; Atwood et al. 2014; Moran Lauter et al. 2014), biotic stress (Kim et al. 2011; Tremblay et al. 2013; Lin et al. 2014; Wong et al. 2014), and throughout soybean development (Zabala et al. 2012; Cho et al. 2013; Jones and Vodkin 2013; Kour et al. 2014).

For many non-model species, RNA-Seq technologies have facilitated the development of a de novo transcriptome sequence prior to, or instead of, the genome sequence. This is a particularly attractive approach for outcrossing species such as alfalfa (Yang et al. 2011; O'Rourke et al. 2015), or those with large, complicated genomes such as lentil, pigeon pea, chickpea, and lupine (Dubey et al. 2011; Kaur et al. 2011; O'Rourke et al. 2013a; Kudapa et al. 2014). For soybean, which has a high-quality genome sequence, the utilization of RNA-Seq reads to develop de novo transcriptomes may seem counterintuitive. However, reference genomes have limitations. Genes may not be present in the genotype used as the reference, but may confer important traits of interest. Developing de novo assemblies for a species with a reference genome sequence can be done by either aligning all reads from the RNA-Seq experiment to each other, developing a new transcriptome, or by mapping the reads to the reference genome then using unaligned reads to assemble novel transcripts not present in the reference. Most de novo assembly programs including Velvet/Oases, Trinity, and SOAPdenovo (Zerbino and Birney 2008; Grabherr et al. 2011; Luo et al. 2012; Schulz et al. 2012) utilize de Brujin graphs to scaffold sequence reads and re-constitute transcript sequences. This usually results in far more sequences than are reasonably expected. Programs to collapse redundant transcripts include CD-Hit (Fu et al. 2012) and CAP3 (Huang and Madan 1999); however, they often result in the loss of alternative splice variants. It is important to ensure a de novo transcriptome accurately reflects the biological reality. The N50 statistic is used to judge the quality of an assembly; the higher the N50, the better the assembly. Programs like DETONATE can assist in determining the best de novo assembly program, optimizing parameters, and determining sequencing support for an assembled contig (Li et al. 2014a). To determine the completeness or coverage of the transcriptome, Gongora-Castillo

and Buell (2013) have proposed the following three parameters: the proportion of assembled reads, the saturation of improvements in N50 contig size, and the number of contigs that can be annotated. These analyses help ensure that data generated by using the assembly is biologically sound and informative.

The regulation of gene expression is complex, and it includes regulating transcription factor expression, DNA methylation, and histone modifications. While functional genomics tools traditionally focus on correlating DNA sequences and phenotypes, methylation patterns also play an important role in phenotypic expression. Methylation patterns are one facet of the rapidly expanding field of epigenetics. Broadly, epigenetics is defined as the study of heritable phenotypic changes not due to changes in the DNA sequence. Thus, epialleles are alleles with the same DNA sequence, but different methylation patterns. There are three types of epialleles: obligate, the methylation pattern is completely dependent on a DNA sequence variant; pure, the epiallele is maintained independent of the DNA sequence; and facilitated, the genetic variant can influence the genetic state, but not as reliably as the obligate epialleles (Schmitz 2014). Much like identification of suites of genes differentially expressed in plants exposed to different conditions, researchers can also identify differentially methylated regions (DMRs), which are heritable over multiple generations. Methylation is usually highest in heterochromatic regions composed of tandem or inverted repeats and transposons (Song et al. 2013)

Methylation patterns have been investigated in soybean since 1994 when researchers performed Southern blots using hypomethylated genomic DNA as probes to investigate the genomic structure and methylation patterns of duplicated regions of the soybean genome (Zhu et al. 1994). In 1999, researchers were able to systematically map the centromeric regions of soybean by comparing the mapping profiles of two amplified fragment length polymorphism (AFLP) markers targeting genomic regions with dissimilar methylation properties (methylation insensitive vs. methylation sensitive) and combining this with known genomic properties (Young et al. 1999). The methylation patterns between the annual wild soybean (*Glycine soja*) lines and cultivated soybean (*G. max*) lines were identified using methylation-sensitive polymorphism (MASP) analysis (Zhong et al. 2009). Both methylation polymorphism types were more frequently observed in *G. max* than in *G. soja*, suggesting methylation polymorphisms have been selected during domestication (Zhong et al. 2009). An AFLP analysis of the same plants revealed no correlation between genetic and methylation polymorphisms, suggesting these two polymorphisms were generated and/or maintained by separate and independent mechanisms (Zhong et al. 2009).

More recently, the genomes, methylomes, and transcriptomes of 83 recombinant inbred lines (RILs) and their parents were sequenced to determine how epialleles impact phenotypic variation (Schmitz et al. 2013). This study found that recently duplicated genomic features were more highly methylated than those arising from earlier duplications. Specifically, analyzing pairs of paralogs revealed that while CG methylation was similar for both pairs, CHG and CHH methylation (where H represents A, C, or T) was statistically enriched in one paralog compared to the other. This corresponds to the differential expression patterns measured between the paralog pairs. Since CHG and CHH methylation are a result of RNA-directed DNA methylation (RdDM), these results suggest this pathway is highly active in the soybean genome. The authors proposed these methylation patterns and subsequent impacts on gene expression are a mechanism utilized by the plant to regulate gene expression through the gene loss or subfunctionalization processes (Schmitz et al. 2013). In a separate study, DNA methylation patterns were identified and compared between roots, stems, leaves, and cotyledons of developing seeds. Between these tissue types, 2162 DMRs (each unique to a single tissue type) were identified (Song et al. 2013). As expected, hypomethylated regions significantly correlated with increased expression of flanking genes, though there was no correlation with downregulation of gene expression near hypermethylated DMRs.

Additionally, the results from this study suggest high CHG and CHH methylation within a soybean gene may induce gene silencing (Song et al. 2013).

Methylation also plays an important role in phenotypic traits. Specific seed coat variations are a result of increased methylation of a transposon, which inhibits the binding of a transposase, thus facilitating mRNA processing and subsequent anthocyanin biosynthesis (Zabala and Vodkin 2014). *RhgI*, which mediates soybean cyst nematode (SCN) resistance, has 8 DMRs between SCN susceptible (containing a single *RhgI* locus repeat) and SCN resistant lines. Interestingly, hypermethylated regions exhibit increased transcript abundance, though expression may be even higher if methylation is reduced. Methylation of *RhgI* adds to the complexity of phenotypic control and was likely important in *RhgI* evolution (Cook et al. 2014). Song et al. (2012) identified ten transcription factors whose expression is methylation dependent and are induced by salinity stress, confirming the critical role of methylation in soybean salinity stress tolerance.

10.3 Forward Genetic Resources for Soybean Improvement

Forward genetics studies, identifying a mutant phenotype then uncovering the genetic sequence causing the phenotype of interest, have been widely adopted in soybean. There are many ways to perturb the functions of plant genes, and mutant populations have been developed through the use of random mutagens including chemicals, radiation, or DNA elements such as transposons and *Agrobacterium* T-DNA. Random mutagenesis is heritable and stable, but it requires intensive screening to identify mutant phenotypes. Following identification of a mutant of interest, the responsible mutation must be fine mapped before it can be cloned and sequenced. Alternatively, mutagenized populations can be indexed and characterized by PCR-based sequencing methods. Individual mutants carrying mutations in the genes of interest can be tested for altered phenotypes. These approaches essentially generate random mutant populations, and then methods are required to de-convolute the mutations in order to identify those that are of interest. Once the population is indexed, it becomes extremely valuable in functional genomics applications. Prime examples of this are *Arabidopsis thaliana* T-DNA lines (Alonso et al. 2003). Alternatively, methods may be developed to enable rapid identification of mutations of interest from populations like targeting induced local lesions in genomes (TILLING) (Cooper et al. 2008).

Soybean has undergone several genetic bottlenecks resulting in low genetic diversity, particularly among elite cultivars (Hyten et al. 2006). While a number of spontaneous mutation events have been identified and characterized in soybean, induced mutagenesis provides a mechanism to increase heritable diversity at a faster rate than seen in nature. Unlike transgenic plants, there are no regulatory restrictions on the handling and transfer of mutant plants, making these ideal to incorporate into breeding programs. Additionally, while transgenic approaches result in few transgenic plants, each mutagenesis event results in a large number of independent mutants. Despite these advantages, mutagenesis does have limitations: Different genotypes are more tractable to mutagenesis than others and even among mutable genotypes, environmental effects can severely impact phenotypes between replicate experiments. Recent years have seen the development and utilization of multiple soybean mutant populations, each with their own advantages, which can be utilized in both forward and reverse genetic studies. In this section, we will present each population type and explore its utility as a functional genomics tool. As soybean is prized for its seeds, it is no surprise that the majority of the studies involving mutant populations focus on altered seed compositions.

10.3.1 Ionizing Radiation

Ionizing radiation includes X-rays, fast neutrons, and gamma rays. In soybean, mutant populations

developed from X-ray and fast neutron radiation are publicly available. Until recently, ionizing radiation was believed to only induce large deletions. Through the adaptation of high-throughput genomics tools including CGH arrays and high-throughput genomic re-sequencing, researchers have determined ionizing radiation induces deletions, duplications, inversions, and translocations of all sizes (Bolon et al. 2011; Belfield et al. 2012; O'Rourke et al. 2013b). Ionizing radiation has been used to develop soybeans with increased yield, earlier flowering times, increased germination rates, decreased dehiscence, increased biotic and abiotic stress resistance, decreased anti-nutritional components, and altered protein and oil compositions.

X-ray mutagenesis was used to develop the earliest mutant populations. In 1955, L.F. Williams identified L67-3483, a tan saddle seed coat mutant, from X-ray irradiated Clark (Rode and Bernard 1975). The locus underlying this trait was subsequently mapped by Kato and Palmer (2003) to the top of molecular linkage group D1b +W (chromosome 2). Additional X-ray irradiation populations have been produced at Saga University in Japan resulting in soybean lines with high oleic acid, high linolenic acid, and high stearic acid (Takagi et al. 1989; Rahman et al. 1994, 1995). Similar to William's research, these early studies provided excellent material for breeding programs, but underlying sequences and genes were not identified. More recently, Anai et al. (2012) have used X-ray irradiated plants to identify two high palmitic acid mutants. Sequence analysis confirmed the mutation of *GmKASIIA* and *GmKASIIB* in two mutants. However, sequence analysis of a third mutant, with a similar phenotype, showed that it was wild type for both *GmKASII*, suggesting additional genes are involved in regulating the palmitic acid content of soybean seeds. X-ray mutagenesis was also employed at the University of Missouri where a forward genetic screen of X-ray mutants identified three mutant lines with increased stearic acid (Gillman et al. 2014). Further genetic analyses, including a reverse genetic screen of an ethylmethane sulfonate (EMS) mutant population, determined the phenotype resulted from mutations in *SACPD-C* (Gillman et al. 2014).

Mutants resulting from gamma irradiation have proven particularly useful in developing plants with improved nutritional quality including reduced phytate levels, Kunitz trypsin inhibitor activity, and lipoxygenase-free seeds (Kim et al. 2010a; Lee et al. 2011, 2014; Yuan et al. 2012). These mutants are easily integrated into large breeding programs. In Japan, almost 10% of the total acreage of soybean is from gamma ray-induced mutants (Nakagawa 2009). Nine of the commercially available cultivars are 'indirect' mutant cultivars (Nakagawa 2009), descended from gamma-irradiated parental lines. Similarly, in China, over 16.33×10^6 ha are planted with soybean varieties derived from gamma irradiation (Kharkwal and Shu 2009). India and Thailand also utilize gamma-irradiated soybeans in their breeding programs. Breeders in India identified over 55 gamma-irradiated mutants with a variety of altered traits, seven of which were publicly released (D'Souza et al. 2009; Kharkwal and Shu 2009).

While the majority of mutant plants are identified and pursued for altered seed properties, multiple groups have used gamma-irradiated populations to identify plants with increased abiotic stress tolerance. Researchers in Thailand have identified mutants with increased resistance to soybean crinkle leaf virus and increased germination rates in extreme dry or wet climates, increasing germination rates from 30 to 70% in the dry season and from 41 to 83% in the wet season (Srisombun et al. 2009). While the specific changes to DNA resulting in flood tolerance have not yet been identified, gene expression analyses of mutant plants grown in flooded and control conditions determined that genes involved in cell wall loosening and proteolysis are not upregulated in the flooded mutant. Additionally, anaerobic metabolism is more efficient in flood-tolerant mutants (Komatsu et al. 2013). All these traits provide interesting avenues for future research on improving abiotic stress tolerance in soybean and other legume species.

Fast neutron (FN) mutagenesis involves bombarding plants with neutrons to induce mutations. Until recently, these mutations were all assumed to be large deletions. Genomics studies have determined that small (<100 base pair) deletions, genome duplications (both large and small), and various translocation and genomic re-arrangements are also induced by FNs (Bolon et al. 2011; Belfield et al. 2012; Anderson et al. 2014). One of the most notable soybean FN mutants is FN37, which lacks the ability to regulate nodule production and thus forms 10× the number of nodules of the wild-type parent (Men et al. 2002). Genetic and physical mapping determined this mutant contained a 460 Kbp deletion on molecular linkage group H (chromosome 12). A dwarf plant resulting from FN mutagenesis was identified by Hwang et al. (2014). Combining forward genetic screens with next-generation sequencing, Hwang et al. (2014) identified an 803 base pair deletion, including part of a peroxidase homolog Glyma15g05831, which likely results in the dwarf phenotype, a beneficial trait in extreme climates. Similarly, high-throughput genomics approaches including CGH arrays, exon capture, and next-generation sequencing have been used to identify and confirm changes in the genome likely underlying changes in plant architecture or seed quality (Bolon et al. 2011, 2014b). The entirety of the data assembled from this FN population including seeds, photographs of obvious phenotypic changes, and NIR data on seed composition has been collected and is publicly available at www.soybase.org/mutants/. Here, sresearchers can search by sample name, genes of interest, trait, images, or phenotypes to identify a suite of mutants for which they can request seed for use in their own research projects.

10.3.2 Insertional Mutagenesis

Insertional mutagenesis has also been used to develop mutant populations. In soybean, individual transformation events are both time and labor intensive. Using transposon and retrotransposon tagging, a multitude of mutations can be derived from a single transformation event. Soybean currently has three transposon tagging mutant populations developed from three unique transposon systems (Chap. 12): *Ac*/*Ds*, *mPING*, and *Tnt1*. The *Ac*/*Ds* transposon system, originally identified by Barbara McClintock in maize, has two major advantages: (1) the preference of these transposons to insert into exonic regions, increasing the likelihood of interrupting gene function and (2) the ability to determine the location of insertion via PCR, based on the known sequence of the transposons. This system has been used to create activation tagging, gene, and enhancer trap element mutants (Mathieu et al. 2009). The miniature inverted-repeat transposable element (*mPING*), originally identified in rice, has a high transposition frequency in soybean, though transposition may be developmentally regulated and plants may continue to experience transposition events, further complicating pheno/genotyping (Hancock et al. 2011). Unlike the *Ac*/*Ds* transposon system which tends to insert near their original point of origin (Mathieu et al. 2009), *mPING* transposes to unlinked genomic positions, usually within 2.5 Kbp of a gene sequence (Hancock et al. 2011). *Tnt1*, a retrotransposon from tobacco, can be induced to transpose in species (including soybean) through tissue culture (Cui et al. 2013). *Tnt1* mutagenesis requires fewer initial transgenic events than the *Ac*/*Ds* system, and the insertions induce stable and heritable changes to the genome. The data associated with all three types of insertional mutants have been deposited at the University of Missouri soybean genome database available at the following Web page digbio.missouri.edu/gmgenedb/index.php. These mutant populations are valuable functional genomics tools for deciphering gene function in soybean and other legumes. Recently, an endogenous CACTA-type transposon *Tgm9* has been identified and utilized in creating a soybean mutant population (Chap. 12).

10.3.3 Chemical Mutagenesis

Chemically mutagenized soybean populations have been developed from both EMS (ethyl methanesulfonate) and NMU (*N*-nitroso-*N* methylurea) mutagenesis. EMS and NMU mutagenesis induce single nucleotide polymorphisms (SNPs) either A to T or G to C transitions. Such mutants are often utilized in reverse genetic studies such as TILLING where DNA from mutagenized plants is screened using primers developed for a specific DNA sequence of interest. The first successful use of TILLING in soybean identified mutations in the omega-6 fatty acid desaturase gene *FAD2-1A*. Further analyses determined these were missense mutations, one of which increased the oleic acid content of seeds, a positive change for soybean seed profiles (Dierking and Bilyeu 2009). Subsequent studies used TILLING in two novel EMS mutant populations identified plants with mutations in the *GmFAD2-1b* genes (Hoshino et al. 2010). Crossing *GmFAD2-1a* and *GmFAD2-1b* mutants produced soybeans with seed oleic acid content >80% (compared to ~18% of normal soybeans) (Hoshino et al. 2010).

In addition to the protein and oil compositions of soybean seeds, improved disease resistance is an important area of research. The SCN is one of the most devastating pathogens affecting soybean production. To identify genes conferring SCN resistance, researchers performed a TILLING screen on an EMS mutagenized soybean population generated from the SCN resistant cultivar 'Forrest' (Liu et al. 2011, 2012). Screening this population for mutations in the obvious LRR-RLK gene identified mutants, but the mutants did not exhibit altered SCN resistance (Liu et al. 2011). However, screening the EMS population for the neighboring serine hydroxymethyltransferase (SHMT) gene identified two mutants with missense mutations, one of which was predicted to be deleterious. Upon SCN infection, the deleterious mutation correlated with the loss of SCN resistance phenotype. Since SHMT is not a canonical resistance gene, additional genetic analyses were employed to confirm its role in SCN resistance. The SHMT gene is involved in the one-carbon metabolism pathway and is conserved across species, making it a unique resistance mechanism.

10.4 Reverse Genetic Resources for Soybean Improvement

Reverse genetic studies start with a gene sequence of interest whose function is investigated utilizing various genetic and genomics tools. In this section, we discuss RNA-silencing techniques that can be used to downregulate or silence the expression of soybean genes. There are several RNA-silencing-based technologies that use double-stranded RNAs (dsRNAs), inverted-repeat RNAs (irRNAs), or miRNAs to knock down the expression of plant genes. DsRNAs are delivered using viral vectors or transgenes, while irRNAs and miRNAs are delivered using transgenes. RNA-silencing-based approaches are not equivalent to null mutations, because it is probable that some of the target gene product(s) is still produced, and there is variability in the extent to which RNA silencing occurs between lines.

Plant RNA-silencing systems can be manipulated, so that plants specifically shut down expression of their own genes. RNA-silencing technologies enable researchers to induce loss-of-function of the targeted gene(s) of interest. This is possible because RNA-silencing systems are programmed by small RNAs that are produced from dsRNA precursors. DsRNAs can be introduced into plant cells in the form of recombinant viruses that replicate through double-stranded RNA intermediates, form secondary structures, or by expression of irRNAs and precursors of miRNAs. Double-stranded viral RNA and irRNAs are cleaved by Dicer-like (DCL) enzymes into siRNAs ranging from 21 to 24 nucleotides in length, depending on the DCL that did the processing. The siRNAs are then bound by argonaute (AGO) proteins, and the RNA-induced silencing complex (RISC) is formed. The siRNAs program AGO to specifically cleave complementary RNA sequences present in the cell. Therefore, if a

siRNA is complementary to a sequence within a plant mRNA, that mRNA will be targeted for degradation, and the expression of that gene will be reduced or possibly abolished. In short, recombinant viruses or inverted-repeat RNAs program RISC to target plant mRNAs by inducing the accumulation of pools of overlapping siRNAs that are complementary to plant mRNAs, resulting in decreased expression of target genes. Gene-silencing strategies that utilize miRNAs also exploit DCL and AGO to process the microRNA precursor, but they result in the production of single, specific miRNAs or the production of phasiRNAs that can act *in trans* on complementary mRNA targets (Schwab et al. 2006; Felippes et al. 2012; Carbonell et al. 2014, 2015).

10.4.1 Virus-Induced Gene Silencing (VIGS)

VIGS is a transient method for reducing the expression of targeted plant genes to create loss-of-function phenotypes (Becker 2013). VIGS relies on the development of infectious clones of viruses that are suitable for gene-silencing applications. The ideal VIGS virus possesses a relatively weak suppressor of RNA silencing, and its genome tolerates and stably maintains foreign inserts over the infection time course in experimental plants. For VIGS applications, the viral genome is modified to contain a cloning site that enables the insertion of fragments of plant genes to be silenced. The recombinant viruses are inoculated into plants using techniques such as biolistics, rub-inoculation, or *Agrobacterium* infiltration. As the recombinant virus spreads systemically throughout the plant, it carries the fragment of the plant gene, which programs the RNA-silencing system to degrade mRNAs that contain complementary sequences. While silencing may persist through the life of the plant or even into the next generation, the technique is considered transient, because the viruses do not insert into the plant genome.

Infectious clones of five viruses have been used for VIGS in soybean with varying degrees of efficacy, and they have been reviewed in detail elsewhere (Whitham et al. 2015). *Bean pod mottle virus* (BPMV) has been used extensively by us and others for VIGS applications, and detailed methods on two different BPMV systems are available (Kachroo and Ghabrial 2012; Zhang et al. 2013). The main difference between the two BPMV systems is how the initial inoculum is generated—one uses in vitro transcription to produce infectious RNA transcripts that are rubbed onto soybean plants (Kachroo and Ghabrial 2012), and the other uses biolistics or rub-inoculation to directly inoculate plants with plasmid DNA carrying the infectious clones (Zhang et al. 2013; Whitham et al. 2016a). After inoculation, BPMV produced from the two systems moves systemically and induces gene silencing to similar extents.

BPMV is a bipartite virus that has two components to its genome, RNA1 and RNA2. Both RNAs contain 5′ and 3′ untranslated regions that flank single, large open reading frames encoding polyproteins that are processed into the individual mature proteins by protease functions encoded by RNA1. RNA1 also encodes replication functions, and the shorter RNA2 encodes the movement and capsid functions. Because the essential viral functionalities are distributed between the two RNAs, they are both necessary for systemic infection. RNA2 is about 2.3 Kbp shorter than RNA1, and therefore, it is RNA2 that has been engineered to accept foreign inserts (Zhang and Ghabrial 2006). The original cloning sites were placed within the open reading frame of RNA2 between the movement protein and the large subunit of the capsid protein (Zhang and Ghabrial 2006; Zhang et al. 2009, 2010). This position requires that any inserted sequence maintains the viral open reading frame, and it has been shown to have great utility for VIGS and for expression of proteins as well. A second position for the cloning site is immediately after the stop codon in the open reading frame of RNA2 (Zhang et al. 2010). This position can only be used for VIGS, and it offers more flexibility in the choice of target sequences and insert orientations because the viral reading frame does not have to be conserved. Using this cloning

position, we have found that gene fragments cloned in the anti-sense orientation are most effective at silencing the target genes (Zhang et al. 2010; Juvale et al. 2012). More recently, a third cloning site in the 5′ untranslated of RNA2 was proposed for the insertion of host gene fragments for VIGS (Ali et al. 2014). Similar to the 3′ untranslated region, fragments inserted into the 5′ untranslated region are also not constrained by presence of stop codons.

BPMV VIGS is used in focused, hypothesis-driven studies to investigate functions of one or a few genes and in large-scale screens of candidate genes identified through approaches such as transcriptome profiling or proteomics (Pandey et al. 2011; Zhang et al. 2012; Cooper et al. 2013). The large-scale screening capability is made possible by the transient and rapid nature of VIGS. Recombinant BPMV clones carrying fragments of plant genes approximately 200–300 base pairs in length are simple to assemble (Whitham et al. 2016a). Systemic virus infections occur within two to three weeks after inoculation, and at this time, silencing of the target gene is established in the symptomatic tissues (Zhang et al. 2010). Because the initial inoculation procedures are somewhat laborious and relatively expensive, we typically store the infected tissue from bombarded plants and use the infected leaf sap as inoculum for the actual experiments that involve many plants, which are rub-inoculated (Zhang et al. 2013; Whitham et al. 2016a). Inserts in the 200–300 base pair size range are relatively stable in BPMV, and so, it is possible to passage the virus once without high concern for them being deleted. However, it is prudent to confirm this. Regardless of the inoculation strategy, it is possible to generate information on the functions of genes within about two months using the BPMV VIGS system.

The BPMV system has been mainly applied to studying traits involved in interactions with biotic and abiotic stresses (Atwood et al. 2014; Liu et al. 2015). VIGS candidate genes used in these studies have been identified based on a number sources including transcriptome profiling, proteomics, homologs from model systems like *Arabidopsis thaliana*, and genetic mapping. There has been a lot of success with identifying genes that affect soybean interactions with foliar pathogens such as soybean rust, downy mildew, *Soybean mosaic virus*, and *Pseudomonas syringae* pv. *glycinea* (Liu et al. 2015; Whitham et al. 2016b). In addition, the BPMV system can be used to silence genes in the roots (Juvale et al. 2012), and protocols have been developed that enable identification of genes involved in soybean–SCN interactions (Liu et al. 2012; Kandoth et al. 2013). With regard to abiotic stresses, the BPMV system has been used to assess functions of genes in iron-deficiency chlorosis and salt stress (Rao et al. 2013; Atwood et al. 2014).

The ability of BPMV to induce silencing in a variety of tissues over the course of plant development (Juvale et al. 2012) suggests it will be useful for a variety of traits besides interactions with the environment. The current bias that exists in the literature is likely due to the research interests of the groups who have developed the systems and their collaborators. BPMV has been shown to silence target genes in leaves, roots, petioles, stems, and flowers. However, more work is needed to investigate extent to which VIGS may be transmitted to the seed. Seed transmission of VIGS in soybean has been demonstrated for the *Apple latent spherical virus* (ALSV) vector in the soybean cultivar 'Enrei' (Yamagishi and Yoshikawa 2009). ALSV has been shown to be useful for identifying genes involved in developmental traits in soybean (Takahashi et al. 2013) as well as both defensive and developmental traits in other plant species (Li and Yoshikawa 2015). Detailed methods in the use of ALSV for VIGS in soybean have been published (Yamagishi and Yoshikawa 2013).

10.4.2 IrRNA

DsRNA can be produced in plant cells via transgenes that are designed to express inverted repeats of fragments of target genes. The inverted repeats may be separated by spacers or introns. In the case of spacers, the transcript forms a double-stranded stem with a loop, which is a structure known as a hairpin. A detailed analysis

has been conducted of silencing of the soybean *FAD3* homologs, *GmFAD3A*, *GmFAD3B*, and *GmFAD3C*, which share 100, 96.5, and 84.3% identity, respectively, with a 318 base pair hairpin RNA construct (Lu et al. 2015). *GmFAD3A* and *GmFAD3B* mRNAs were silenced very effectively, but *GmFAD3C* mRNA was silenced to a lesser extent. The reduced level of *GmFAD3C* mRNA silencing correlated with the observation that *GmFAD3C* shared perfect complementarity with relatively few of the siRNAs produced by the hairpin transgene, and this in turn led to inefficient cleavage of the mRNA. These data highlight the challenges and limitations of using RNA-silencing approaches to target gene family members in soybean.

In a true functional genomics application of silencing using inverted-repeat transgenes, Danzer et al. (2015) silenced 53 soybean transcription factors that were selected based on detailed transcriptome analyses of soybean seed development. Transgenic soybean lines were made to express an irRNAs targeting a 150–200 base pair fragment in each transcription factor gene. The RNA-silencing lines were screened for defects in seed development and vegetative growth, and mutant phenotypes were observed for silencing of three of the transcription factors. The functions of one of the transcription factors, *SPEECHLESS* (Glyma04g41710), were investigated in depth. The RNA-silencing construct designed against Glyma04g41710 was effective at silencing it and the other three paralogs that shared 98, 85, and 83% nucleotide identity. Unlike the case with *FAD3*, mRNA transcripts of all the *SPEECHLESS* paralogs were silenced to similar levels. The mutant phenotype correlated with decreased mRNA transcript accumulation, and it was similar to Arabidopsis *speechless* mutants based on defects in stomata development. The availability of Arabidopsis *speechless* mutants made it possible to perform a transgene complementation study that confirmed that the soybean *SPEECHLESS* homolog in question was indeed a functional ortholog. This study illustrates the power of utilizing resources, like Arabidopsis mutants, for determining the functions of soybean genes, which can be coupled with gene expression and RNA-silencing studies.

10.4.3 Applications for MiRNA in Soybean Functional Genomics

MiRNAs are small RNAs like the siRNAs, but their biogenesis within the cell occurs via a distinct pathway with processing mediated by DCL1. MiRNAs direct AGO1 to cleave mRNA transcripts containing complementary sequences. MiRNAs can be artificially designed to specifically target plant genes by exchanging the sequence of a natural miRNA with an artificial one (Schwab et al. 2006). Expression of the artificial miRNA precursor can be placed under control of any desired promoter. In soybean, artificial miRNAs have been expressed under the control of the constitutive polyubiquitin promoter (*Gmubi*) and the soybean 7S globulin promoter, which drives seed-specific expression. Melito et al. (2010) used artificial miRNAs in hairy root assays to demonstrate that a leucine-rich repeat receptor kinase was not a candidate for the SCN resistance gene, *Rhg1*. Artificial miRNAs have also been expressed in a seed-specific manner in stably transformed plants to investigate the effects of 7S globulin proteins on seed nitrogen sources and total protein content (Yamada et al. 2014). The 'P-SAMS amiRNA Designer' is a web-based tool that can be used to design artificial miRNAs with novel targets for soybean and other plant species (Carbonell et al. 2014, 2015).

MiRNAs of the 22 nucleotide size class can also be used to trigger the accumulation of siRNAs that are loaded onto AGO proteins and subsequently silence expression of a target gene (s). These siRNAs are known as phasiRNAs, and if they are demonstrated to direct cleavage of a target RNA *in trans*, then they are called trans-acting siRNAs (tasiRNAs). In an approach that has been referred to as miRNA-induced gene silencing (MIGS) (Felippes et al. 2012), the 22 nucleotide miRNA target sequence is placed upstream of a fragment of a target gene, which

results in the production of synthetic tasiRNAs (syn-tasiRNAs) that direct silencing of the target gene. In soybean, this approach has been demonstrated to silence genes in both hairy roots and transgenic plants (Jacobs et al. 2016). An advantage of MIGS over inverted repeats is that it is easier to construct the clones, because only a single copy of the target fragment in a single orientation is required. However, it is necessary to first identify a miRNA that initiates production of tasiRNAs in the appropriate tissues or cell types under the desired conditions or co-express a miRNA that is known to induce tasiRNA biogenesis. The P-SAMS syn-tasiRNA Designer Wizard is a web-based tool that can be used to design syn-tasiRNAs for soybean and other plant species (Carbonell et al. 2014).

10.5 Bridging the Genotype-to-Phenotype Gap—Rapidly Developing Areas that Will Accelerate Discovery of Soybean Gene Function

10.5.1 Using Next-Generation Sequencing to Characterize Complex Traits

A high-quality reference genome has enabled researchers to begin to examine complex traits, such as yield, domestication, and drought tolerance. Re-sequencing data from different soybean accessions can be assembled to the existing Williams 82 reference genome. In 2010, Kim et al. (2010b) assembled re-sequencing data from undomesticated soybean (*G. soja*) to the reference. This identified 712 genes that were partially or completely lost in *G. max*, likely during the process of domestication. Lam et al. (2010) re-sequenced 17 wild and 14 cultivated soybean genomes revealing the phylogenetic relationships between lines, identifying linkage disequilibrium blocks and generating over 200,000 SNPs that could be used for mapping and association studies. Li et al. (2013) leveraged the Lam et al. (2010) re-sequencing data with additional re-sequencing data from 25 diverse soybean accessions. The 50 Chinese lines represented wild soybean, land races, and elite cultivars allowing the researchers to differentiate between candidate genes associated with soybean domestication and improvement. Xu et al. (2013) used genome re-sequencing approaches to generate and validate SNPs that were used for QTL mapping of root-knot nematode resistance. Their approach identified two candidate genes responsible for resistance. Chung et al. (2014) used re-sequencing data of ten cultivated and six wild Korean accessions to identify 206 candidate domestication regions with lower diversity than observed within the wild accessions. Some of these regions contained genes with homology to known domestication genes identified in other plant species, while others appear to be novel in soybean. In 2015, Zhou et al. (2015) re-sequenced 302 wild and cultivated accessions allowing the identification of selective sweeps and copy number variants between genomes. In addition, genome-wide association studies were used to identify genomic regions associated with oil content, plant height, and pubescence. Bolon et al. (2014a) used genome re-sequencing to characterize FN radiation mutants.

Rapid decreases in the cost of next-generation sequencing have also facilitated the adoption of de novo sequencing strategies for soybean. In 2014, Li et al. (2014b) used next-generation sequencing of seven phylogenetically and geographically distinct *G. soja* accessions to characterize the *G. soja* pan genome. Eighty percent of the pan genome was present in all seven genomes, representing the core *G. soja* genome. Twenty percent of the pan genome was considered dispensable, with dispensable genes showing greater levels of sequence diversity. Core genes were enriched for biological processes including growth and reproduction. Qi et al. (2014) combined de novo assembly of a wild soybean accession with re-sequencing of a recombinant inbred population to identify the *GmCHX1* salt tolerance gene. Collectively, these studies demonstrate the ability that researchers now have to target complex traits for soybean improvement.

Genes identified using next-generation sequencing base approaches are key targets for characterization using cutting-edge functional genomics tools.

10.5.2 Epigenetics

Soybean epigenetic studies are an exciting area of future research, adding another level to the potential variation resulting in phenotypic changes. It is well recognized that soybean underwent a genetic bottleneck, severely limiting genetic diversity. Methylation studies in soybean, and other species, have shown a large number of allelic variants are silenced by methylation, thus altering methylation patterns may be a mechanism to generate phenotypic diversity (Schmitz 2014). As the majority of epialleles complied with Mendelian inheritance, methylQTLs can be identified for >90% of DMRs (Schmitz et al. 2013). Identifying methylQTLs and the causal epialleles may be a major challenge in soybean genetic studies. Epialleles induced by environmental conditions are a huge area of interest as understanding and identifying these may aide in understanding local adaptation and developing soybeans for extreme climates and climate changes. Finally, generating specific epialleles to test epiQTLs, induce desirable phenotypes, or regulate gene expression in response to stress, all without changing the underlying DNA sequence, has the potential to revolutionize soybean breeding and functional genomics studies.

10.5.3 Genome Editing

Precise genome editing involves the use of technologies that enable researchers to specifically alter the genetic code in a heritable manner without requiring a transgene or even making a transgenic plant (Luo et al. 2015). This sets genome editing apart from the RNA-silencing approaches discussed earlier and from standard methods used to overexpress genes in plants. Genome editing is made possible through the development of site-directed nucleases that can specifically recognize a DNA sequence and cleave it (Baltes and Voytas 2015). Meganucleases, zinc finger nucleases (ZFNs), and transcription activator-like effector nucleases (TALENs) provide site selectivity by protein–DNA interactions. Clustered regularly interspaced short palindromic repeats/Cas9 (CRISPR/Cas9) systems confer site specificity by complementarity between guide RNA and the DNA target site. The ZFNs and TALENs are modular synthetic proteins that fuse the DNA recognition domain with the *Fok*I restriction enzyme, which functions as a dimer. Therefore, ZFNs and TALENs are designed in such a way that it is necessary to express two proteins that have their recognition sites flanking the cleavage site. These nucleases only cleave when both recognition sites are bound and the *Fok*I enzyme is able to dimerize and cleave DNA, which ensures specificity. In the CRISPR/Cas9 system, the guide RNA mediates the target site recognition and DNA cleaved by nuclease activity of the Cas9 protein. Therefore, the CRISPR/Cas9 system is simpler and very easy to program by simply supplying different guide RNAs. Xie et al. (2014) estimated than more than 90% of the soybean transcriptome could be targeted using the CRISPR/Cas9 system.

Cleavage of the DNA target by these nucleases results in a double-strand break that can be repaired by one of two pathways, non-homologous end joining (NHEJ) and homologous recombination (HR) (Baltes and Voytas 2015). NHEJ results in insertions, deletions, and point mutations due to imperfect repair of the double-strand break site. NHEJ, therefore, results in targeted mutagenesis, which is useful for making loss-of-function alleles that are valuable in functional genomics studies. HR involves the double-strand break plus a donor DNA molecule that carries the sequence to insert or modify. The donor DNA is flanked by sequence that is identical to either side of the double-strand break enabling it to be inserted through HR. HR is valuable for crop bioengineering, because it allows for the precise replacement of sequences with the desired ones. However, accomplishing HR is more

technically challenging, because both the site-directed nuclease and the donor template must be introduced into the plant to create the specific edit.

The ZFN, TALEN, and CRISPR/Cas9 systems have been used for mutagenesis via NHEJ in soybean (Curtin et al. 2011, 2013; Haun et al. 2014; Jacobs et al. 2015; Sun et al. 2015), and the CRISPR/Cas9 system was used to mediate allele replacement via HR (Li et al. 2015). A general approach is to design the site-directed nuclease and then test its ability to induce mutations at the target locus or loci in hairy roots. Constructs that induce mutations efficiently are then used in soybean transformation. The hairy root step eliminates unproductive site-directed nuclease constructs and saves significant time and resources needed to produce transgenic soybean lines. Mutations in a variety of loci including transgenes and endogenous genes encoding proteins and non-coding RNAs have been reported. The resulting mutations are most frequently small deletions; however, SNPs and insertions occur at significantly lower frequencies. Mutations can occur in a single allele or in both alleles of a gene. Bi-allelic mutations are advantageous because they allow mutant phenotypes to be observed in hairy roots or directly in primary transgenic plants (Jacobs et al. 2015). Even if mutations are heterozygous, it is possible to obtain homozygous, null segregant (lacking the site-directed nuclease transgene) mutants after only a single generation (Haun et al. 2014).

Duplication of the soybean genome complicates functional analysis of genes because a majority of genes have homoeologs that are nearly identical in nucleic acid sequence. The site-directed nucleases offer the ability to create mutations in one or more genes simultaneously (Curtin et al. 2011; Haun et al. 2014; Jacobs et al. 2015; Sun et al. 2015). In regenerated plants, the resulting mutations can be combined or separated via segregation, enabling the effects of the genes to be studied in isolation or combination. Curtin et al. (2011) used a single ZFN pair to create mutations in soybean *DCL4* paralogs (*DCL4a* and *DCL4b*), although only the *dcl4b* mutant produced viable progeny that segregated for the mutation. A single TALEN pair was used by Haun et al. (2014) to target the two fatty acid desaturase 2 genes (*FAD2-1A* and *FAD2-1B*). Four plants in their study carried mutations in both genes, and they were shown to segregate in the progeny of the original transgenic parent. The CRISPR/Cas9 system was also shown to induce mutations in paralogs by using different guide RNAs separately or in combination (Jacobs et al. 2015). Alternatively, the CRISPR/Cas9 system can tolerate mismatches between the guide RNA and its target, so it is also possible to obtain mutations in paralogs through off target activity (Sun et al. 2015). Finally, the site-directed nucleases result in a spectrum of mutations, which opens the possibility for generating allelic series in genes that could be very useful for understanding their functions (Haun et al. 2014).

10.6 Conclusions

Soybean has become an attractive model system for crops in which it is feasible to conduct functional genomics geared toward understanding plant biology, evolution, and domestication. Functional genomics tools and resources are ever expanding, and new technologies are rapidly adopted by the research community to accelerate discoveries related to the function and regulation of soybean genes. A variety of datasets such as mRNA transcript abundance, protein abundance, short RNA populations, and methylation status provide unprecedented insight into gene regulation during developmental programs and responses to the environment. The datasets coupled with sequences of genetically diverse germplasm, expanding mutant collections, RNA-silencing approaches, and precise genome editing provide a fertile ground for generating and testing hypotheses designed to bridge the genotype-to-phenotype gap.

Acknowledgements This work was supported by USDA-ARS Project 3625-21220-005-00D, the Iowa Soybean Association, the United Soybean Board Project Number 1204, USDA NIFA Hatch Project 3808, and State of Iowa funds. The USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this article is solely for the purpose of

providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

References

Ali AK, Lin J, Han J, Ibrahim KM, Jarjees MM, Qu F (2014) The 5' untranslated region of *Bean pod mottle virus* RNA2 tolerates unusually large deletions or insertions. Virus Res 179:247–250

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseeuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653–657

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

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106

Anderson JE, Kantar MB, Kono TY, Fu F, Stec AO, Song Q, Cregan PB, Specht JE, Diers BW, Cannon SB, McHale LK, Stupar RM (2014) A roadmap for functional structural variants in the soybean genome. Genes Genomes Genet 4:1307–1318

Arikit S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, Valdes-Lopez O, Prince S, Musket TA, Nguyen HT, Stacey G, Meyers BC (2014) An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. Plant Cell 26:4584–4601

Atwood SE, O'Rourke JA, Peiffer GA, Yin T, Majumder M, Zhang C, Cianzio SR, Hill JH, Cook D, Whitham SA, Shoemaker RC, Graham MA (2014) Replication protein A subunit 3 and the iron efficiency response in soybean. Plant Cell Environ 37:213–234

Baltes NJ, Voytas DF (2015) Enabling plant synthetic biology through genome engineering. Trends Biotechnol 33:120–131

Becker A (ed) (2013) Virus-induced gene silencing: methods and protocols. Methods in Molecular biology, vol 975. Springer, Berlin

Belfield EJ, Gan X, Mithani A, Brown C, Jiang C, Franklin K, Alvey E, Wibowo A, Jung M, Bailey K, Kalwani S, Ragoussis J, Mott R, Harberd NP (2012) Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. Genome Res 22:1306–1315

Bolon YT, Haun WJ, Xu WW, Grant D, Stacey MG, Nelson RT, Gerhardt DJ, Jeddeloh JA, Stacey G, Muehlbauer GJ, Orf JH, Naeve SL, Stupar RM, Vance CP (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant Physiol 156:240–253

Bolon Y-T, Stec AO, Michno J-M, Roessler J, Bhaskar PB, Ries L, Dobbels AA, Campbell BW, Young NP, Anderson JE, Grant DM, Orf JH, Naeve SL, Muehlbauer GJ, Vance CP, Stupar RM (2014a) Genome resilience and prevalence of segmental duplications following fast neutron irradiation of soybean. Genetics 198:967–981

Bolon YT, Stec AO, Michno JM, Roessler J, Bhaskar PB, Ries L, Dobbels AA, Campbell BW, Young NP, Anderson JE, Grant DM, Orf JH, Naeve SL, Muehlbauer GJ, Vance CP, Stupar RM (2014b) Genome resilience and prevalence of segmental duplications following fast neutron irradiation of soybean. Genetics 198:967–981

Cannon SB, Crow JA, Grant D (2012) SoyBase and the legume information system: accessing inormation about soybean and other legume genomes. In: Wilson RF (ed) Designing soybeans for 21st century markets. American Oil Chemists' Society, Urbana, pp 49–62

Carbonell A, Takeda A, Fahlgren N, Johnson SC, Cuperus JT, Carrington JC (2014) New generation of artificial MicroRNA and synthetic trans-acting small interfering RNA vectors for efficient gene silencing in *Arabidopsis*. Plant Physiol 165:15–29

Carbonell A, Fahlgren N, Mitchell S, Cox KL Jr, Reilly KC, Mockler TC, Carrington JC (2015) Highly specific gene silencing in a monocot species by artificial microRNAs derived from chimeric miRNA precursors. Plant J 82:1061–1075

Cho YB, Jones SI, Vodkin LO (2013) The transition from primary siRNAs to amplified secondary siRNAs that regulate chalcone synthase during development of *Glycine max* seed coats. PLoS ONE 8:e76954

Chung W-H, Jeong N, Kim J, Lee WK, Lee Y-G, Lee S-H, Yoon W, Kim J-H, Choi I-Y, Choi H-K, Moon J-K, Kim N, Jeong S-C (2014) Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. DNA Res 21:153–167

Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, Bent AF (2014) Distinct copy number, coding sequence, and locus methylation patterns underlie *Rgh1*-mediated soybean resistance to soybean cyst nematode. Plant Physiol 165:630–647

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 8:9

Cooper B, Campbell KB, McMahon MB, Luster DG (2013) Disruption of *Rpp1*-mediated soybean rust immunity by virus-induced gene silencing. Plant Signal Behav 8:e27543

Cui Y, Barampuram S, Stacey MG, Hancock CN, Findley SS, Mathieu M, Zhang Z, Parrott WA, Stacey G (2013) *Tnt1* retrotransposon mutagenesis: a tool for soybean functional genomics. Plant Physiol 161:36–47

Curtin SJ, Zhang F, Sander JD, Haun WJ, Starker C, Baltes NJ, Reyon D, Dahlborg EJ, Goodwin MJ, Coffman AP, Dobbs D, Joung JK, Voytas DF, Stupar RM (2011) Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiol 156:466–473

Curtin SJ, Anderson JE, Starker CG, Baltes NJ, Mani D, Voytas DF, Stupar RM (2013) Targeted mutagenesis for functional analysis of gene duplication in legumes. Methods Mol Biol 1069:25–42

Danzer J, Mellott E, Bui AQ, Le BH, Martin P, Hashimoto M, Perez-Lesher J, Chen M, Pelletier JM, Somers DA, Goldberg RB, Harada JJ (2015) Down-regulating the expression of 53 soybean transcription factor genes uncovers a role for SPEECHLESS in initiating stomatal cell lineages during embryo development. Plant Physiol 168:1025–1035

Dierking EC, Bilyeu KD (2009) New sources of soybean seed meal and oil composition traits identified through TILLING. BMC Plant Biol 9:89

D'Souza SF, Reddy KS, Badigannavar AM, Manjaya JG, Jambhulkar SJ (2009) Mutation breeding in oilseeds and grain legumes in India: accomplishments and socio-economic impact. In: Shu Q (ed) Induced plant mutations in the genomics era. Food and Agriculture Organization of the United Nations and the International Atomic Energy Agency, Rome, pp 55–57

Dubey A, Farmer A, Schlueter J, Cannon SB, Abernathy B, Tutej R, Woodward J, Shah T, Mulasmanovic B, Kudapa H, Raju NL, Gothalwal R, Pande S, Xiao Y, Town CD, Singh NK, May GD, Jackson S, Varshney RK (2011) Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeonpea (*Cajanus cajan* L.). DNA Res 18:153–164

Felippes FF, Wang JW, Weigel D (2012) MIGS: miRNA-induced gene silencing. Plant J 70:541–547

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152

Gillman JD, Stacey MG, Cui Y, Berg HR, Stacey G (2014) Deletions of the *SACPD-C* locus elevate seed stearic acid levels but also result in fatty acid and morphological alterations in nitrogen fixing nodules. BMC Plant Biol 14:143

Gongora-Castillo E, Buell CR (2013) Bioinformatics challenges in de novo transcriptome assembly using short read sequenes in the absense of a reference genome sequence. Nat Prod Rep 30:490–500

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2011) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit A, Adiconis X, Fan L, Raychowdhury R, Zheng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 15:644–652

Hancock CN, Zhang F, Floyd K, Richardson AO, LaFayette P, Tucker D, Wessler SR, Parrott WA (2011) The rice miniature inverted repeat transposable element *mPing* is an effective insertional mutagen in soybean. Plant Physiol 157:552–562

Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP, Stupar RM (2011) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol 155:645–655

Haun W, Coffman A, Clasen BM, Demorest ZL, Lowy A, Ray E, Retterath A, Stoddard T, Juillerat A, Cedrone F, Mathis L, Voytas DF, Zhang F (2014) Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnol J 12:934–940

Hoshino T, Takagi Y, Anai T (2010) Novel *GmFAD2-1b* mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seeds in combination with *GmFAD2-1a* mutant alleles. Breed Sci 60:419–425

Hossain MS, Joshi T, Stacey G (2015) System approaches to study root hairs as a single cell plant model: current status and future perspectives. Front Plant Sci 6:363

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Hwang WJ, Kim MY, Kang YJ, Shim S, Stacey MG, Stacey G, Lee SH (2014) Genome-wide analysis of mutations in a dwarf soybean mutant induced by fast neutron bombardment. Euphytica 203:399–408

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103:16666–16671

Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA (2015) Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnol 15:16

Jacobs TB, Lawler NJ, LaFayette PR, Vodkin LO, Parrott WA (2016) Simple gene silencing using the trans-acting siRNA pathway. Plant Biotechnol J 14:117–127

Jones SI, Vodkin LO (2013) Using RNA-seq to profile soybean seed development from fertilization to maturity. PLoS ONE 8:e59270

Joshi T, Patil K, Fitzpatrick MR, Franklin LD, Yao Q, Cook JR, Wang Z, Libault M, Brechenmacher L, Valliyodan B, Wu X, Cheng J, Stacey G, Nguyen HT, Xu D (2012) Soybean Knowledge Base (SoyKB): a web resource for soybean translational genomics. BMC Genom 13:S15

Juvale PS, Hewezi T, Zhang C, Kandoth PK, Mitchum MG, Hill JH, Whitham SA, Baum TJ (2012) Temporal and spatial *Bean pod mottle virus*-induced gene silencing in soybean. Mol Plant Pathol 13:1140–1148

Kachroo A, Ghabrial S (2012) Virus-induced gene silencing in soybean. Methods Mol Biol 894:287–297

Kandoth PK, Heinz R, Yeckel G, Gross NW, Juvale PS, Hill J, Whitham SA, Baum TJ, Mitchum MG (2013) A virus-induced gene silencing method to study soybean cyst nematode parasitism in *Glycine max*. BMC Res Notes 6:255

Kato KK, Palmer RG (2003) Genetic identification of a female partial-sterile mutant in soybean. Genome 46:128–134

Kaur S, Cogan NO, Pembleton LW, Shinozuka M, Savin KW, Materne M, Forster JW (2011) Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genom 12:265–276

Kharkwal MC, Shu QY (2009) The role of induced mutations in wold food securities. Paper presented at the induced plant mutations in the genomics era, Vienna, Austria

Kim DS, Lee KJ, Kim JB, Kim SH, Song JY, Seo YW, Lee BM, Kang SY (2010a) Identification of Kunitz trypsin inhibitors mutations using SNAP markers in soybean mutant lines. Theor Appl Genet 121:751–760

Kim MY, Lee S, Van K, Kim T-H, Jeong S-C, Choi I-Y, Kim D-S, Lee Y-S, Park D, Ma J, Kim W-Y, Kim B-C, Park S, Lee K-A, Kim DH, Kim KH, Shin JH, Jang YE, Kim KD, Liu WX, Chaisan T, Kang YJ, Lee Y-H, Kim K-H, Moon J-K, Schmutz J, Jackson SA, Bhak J, Lee S-H (2010b) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proc Natl Acad Sci USA 107:22032–22037

Kim KH, Kang YJ, Kim DH, Yoon MY, Moon J-K, Kim MY, Van K, Lee S-H (2011) RNA-seq analysis of a soybean near-isogenic line carrying bacterial leaf pustule-resistant and -susceptible alleles. DNA Res 18:483–497

Komatsu S, Nanjo Y, Nishimura M (2013) Proteomic analysis of the flooding tolerance mechanism in mutant soybean. J Proteom 79:231–250

Kour A, Boone AM, Vodkin LO (2014) RNA-seq profiling of a defective seed coat mutation in *Glycine max* reveals differential expression of proline-rich and other cell wall protein transcripts. PLoS ONE 9: e96342

Kudapa H, Azam S, Sharpe AG, Taran B, Li R, Deonovic B, Cameron C, Farmer AD, Cannon SB, Varshney RK (2014) Comprehensive transcriptome assembly of chickpea (*Cicer arietinum* L.) using Sanger and next generation sequencing platforms: development and applications. PLoS ONE 9:e86039

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

Lee KJ, Kim JB, Kim SH, Ha BK, Lee BM, Kang SY, Kim DS (2011) Alteration of seed storage protein composition in soybean [*Glycine max* (L.) Merrill] mutant lines induced by γ-irradiation mutagenesis. J Agric Food Chem 59:12405–12410

Lee K, Hwant JE, Velusamy V, Ha BK, Kim JB, Kim SH, Ahn JW, Kang SY, Kim DS (2014) Selection and molecular characerization of a lipoxygenase-free soybean mutant line induced by gamma irradiation. Theor Appl Genet 127:2405–2413

Li C, Yoshikawa N (2015) Virus-induced gene silencing of *N* gene in tobacco by *Apple latent spherical virus* vectors. Methods Mol Biol 1236:229–240

Li Y-H, Zhao S-C, Ma J-X, Li D, Yan L, Li J, Qi X-T, Guo X-S, Zhang L, He W-M, Chang R-Z, Liang Q-S, Guo Y, Ye C, Wang X-B, Tao Y, Guan R-X, Wang J-Y, Liu Y-L, Jin L-G, Zhang X-Q, Liu Z-X, Zhang L-J, Chen J, Wang K-J, Nielsen R, Li R-Q, Chen P-Y, Li W-B, Reif JC, Purugganan M, Wang J, Zhang M-C, Wang J, Qiu L-J (2013) Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. BMC Genom 14:579

Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R, Dewey CN (2014a) Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome Biol 15:553

Li Y-H, Zhou G, Ma J, Jiang W, Jin L-G, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang S-S, Zuo Q, Shi X-H, Y-f Li, Zhang W-K, Hu Y, Kong G, Hong H-L, Tan B, Song J, Liu Z-X, Wang Y, Ruan H, Yeung CKL, Liu J, Wang H, Zhang L-J, Guan R-X, Wang K-J, Li W-B, Chen S-Y, Chang R-Z, Jiang Z, Jackson SA, Li R, Qiu L-J (2014b) de novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol 32:1045–1052

Li Z, Liu ZB, Xing A, Moon BP, Koellhoffer JP, Huang L, Ward RT, Clifton E, Falco SC, Cigan AM (2015) Cas9-guide RNA directed genome editing in soybean. Plant Physiol 169:960–970

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

Lin F, Zhao M, Baumann DD, Ping J, Sun L, Liu Y, Zhang B, Tang Z, Hughes E, Doerge RW, Hughes TJ, Ma J (2014) Molecular response to the pathogen *Phytophthora sojae* among ten soybean near isogenic lines revealed by comparative transcriptomics. BMC Genom 15:18

Liu X, Liu S, Jamai A, Bendahmane A, Lightfoot DA, Mitchum MG, Meksem K (2011) Soybean cyst nematode resistance in soybean is independent of the *Rgh4* locus *LRR-RLK* gene. Funct Integr Genom 11:539–549

Liu S, Kandoth PK, Warren SD, Yeckel G, Heinz R, Alden J, Yang C, Jamai A, El-Mellouki T, Juvale PS,

Hill J, Baum TJ, Cianzio S, Whitham SA, Korkin D, Mitchum MG, Meksem K (2012) A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. Nature 492:256–260

Liu JZ, Graham MA, Pedley KF, Whitham SA (2015) Gaining insight into soybean defense responses using functional genomics approaches. Brief Funct Genom 14:283–290

Lu S, Yin X, Spollen W, Zhang N, Xu D, Schoelz J, Bilyeu K, Zhang ZJ (2015) Analysis of the siRNA-mediated gene silencing process targeting three homologous genes controlling soybean seed oil quality. PLoS ONE 10:e0129010

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18

Luo S, Li J, Stoddard TJ, Baltes NJ, Demorest ZL, Clasen BM, Coffman A, Retterath A, Mathis L, Voytas DF, Zhang F (2015) Non-transgenic plant genome editing using purified sequence-specific nucleases. Mol Plant 8:1425–1427

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 (*Glycine max* Merr. L) transposon-based mutagenesis repository. Planta 229:279–289

McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. Nucleic Acids Res 40:4288–4297

McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol 159:1295–1308

Melito S, Heuberger AL, Cook D, Diers BW, MacGuidwin AE, Bent AF (2010) A nematode demographics assay in transgenic roots reveals no significant impacts of the *Rhg1* locus LRR-Kinase on soybean cyst nematode resistance. BMC Plant Biol 10:104

Men AE, Laniya TS, Searle IR, Iturbe-Ormaetxe I, Gresshoff I, Jiang Q, Carroll BJ, Gresshoff PM (2002) Fast neutron mutagenesis of soybean (*Glycine soja* L.) produces a supernodulating mutant containing a large deletion in linkage group H. Genome Lett 3:147–155

Meyer JDF, Silva DCG, Yang C, Pedley KF, Zhang C, van de Mortel M, Hill JH, Shoemaker RC, Abdelnoor RV, Whitham SA, Graham MA (2009) Identification and analyses of candidate genes for *Rpp4*-mediated resistance to Asian soybean rust in soybean. Plant Physiol 150:295–307

Moran Lauter AN, Peiffer GA, Yin T, Whitham SA, Cook D, Shoemaker RC, Graham MA (2014) Identification of candidate genes involved in early iron deficiency chlorosis signaling in soybean (*Glycine max*) roots and leaves. BMC Genom 15:702

Nakagawa H (2009) Induced mutations in plant breeding and biological researches in Japan. Paper presented at the induced plant mutations in the genomics era, Vienna, Austria

O'Rourke J, Yang SS, Miller SS, Bucciarelli B, Liu J, Rydeen A, Bozsoki Z, Uhde-Stone C, Tu ZJ, Allan D, Gronwald JW, Vance CP (2013a) An RNA-seq transcriptome analysis of orthophosphate deficient white lupin reveals novel insights into phosphorus acclimation in plants. Plant Physiol 161:705–724

O'Rourke JA, Iniguez LP, Bucciarelli B, Roessler J, Schmutz J, McClean PE, Jackson SA, Hernandez G, Graham MA, Stupar RM, Vance CP (2013b) A re-sequencing based assessment of genomic heterogeneity and fast neutron-induced deletions in a common bean cultivar. Front Plant Sci 4:210

O'Rourke JA, Fu F, Bucciarelli B, Yang SS, Samac DA, Lamb JF, Monteros MJ, Graham MA, Gronwald JW, Krom N, Li J, Dai X, Zhao PX, Vance CP (2015) The *Medicago sativa* gene index 1.2: a web-accessible gene expression atlas for investigating expression differences between *Medicago sativa* subspecies. BMC Genom 16:502

Pandey AK, Yang C, Zhang C, Graham MA, Horstman HD, Lee Y, Zabotina OA, Hill JH, Pedley K, Whitham SA (2011) Functional analysis of the Asian soybean rust resistance pathway mediated by *Rpp2*. Mol Plant Microbe Interact 24:194–206

Peiffer GA, King KE, Severin AJ, May GD, Cianzio SR, Lin SF, Lauter NC, Shoemaker RC (2012) Identification of candidate genes underlying an iron efficiency quantitative trait locus in soybean. Plant Physiol 158:1745–1754

Qi X, Li M-W, Xie M, Liu X, Ni M, Shao G, Song C, Kay-Yuen Yim A, Tao Y, Wong F-L, Isobe S, Wong C-F, Wong K-S, Xu C, Li C, Wang Y, Guan R, Sun F, Fan G, Xiao Z, Zhou F, Phang T-H, Liu X, Tong S-W, Chan T-F, Yiu S-M, Tabata S, Wang J, Xu X, Lam H-M (2014) Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nat Commun 5:4340

Rahman SM, Takagi Y, Kubota K, Miyamoto K, Kawakita T (1994) High oleic acid mutant in soybean induced by X-ray irradiation. Biosci Biotechnol Biochem 58:1070–1072

Rahman SM, Takagi Y, Miyamoto K, Kawakita T (1995) High stearic acid soybean mutant induced by X-ray irradiation. Biosci Biotechnol Biochem 59:922–923

Rao SS, El-Habbak MH, Havens WM, Singh A, Zheng D, Vaughn L, Haudenshield JS, Hartman GL, Korban SS, Ghabrial SA (2013) Overexpression of *GmCaM4* in soybean enhances resistance to pathogens and tolerance to salt stress. Mol Plant Pathol 15:145–160

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:1

Rode NW, Bernard RL (1975) Inheritance of a tan saddle mutant. Soybean Genet Newsl 2:39–42

Schmitz RJ (2014) The secret garden—epigenetic alleles underlie complex traits. Science 343:1082–1083

Schmitz RJ, He Y, Valdez-Lopez O, Khan SM, Joshi T, Urich MA, Nery JR, Diers B, Xu D, Stacey G, Ecker JR (2013) Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res 23:1633–1674

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092

Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D (2006) Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. Plant Cell 18:1121–1133

Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA, Cannon SB, May GD, Vance CP, Shoemaker RC (2010) RNA-seq atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol 10:160

Song Y, Ji D, Li S, Weng P, Li Q, Xiang F (2012) The dynamic changes of DNA methylation and histone modifications of salt responsive transcription factor genes in soybean. PLoS ONE 7:e41274

Song QX, Lu X, Li QT, Chen H, Hu XY, Ma B, Zhang WK, Chen SY, Zhang JS (2013) Genome-wide analysis of DNA methylation in soybean. Mol Plant 6:1961–1974

Srisombun S, Srinives P, Yathaputanon C, Dangpradub S, Malipan A, Srithongchai W, Kumsueb B, Tepjun V, Ngampongsai S, Kornthong A, Impithuksa S (2009) Achievements of grain legume variety improvement using induced mutation of the IAEA/RAS/5/040 project in Thailand. Paper presented at the induced plant mutations in the genomics era, Vienna, Austria

Sun X, Hu Z, Chen R, Jiang Q, Song G, Zhang H, Xi Y (2015) Targeted mutagenesis in soybean using the CRISPR-Cas9 system. Sci Rep 5:10342

Takagi Y, Hossain ABMM, Yanagita T, Kusaba S (1989) High linolenic acid mutant in soybean induced by X-ray irradiation. Jpn J Breed 39:403–409

Takahashi R, Yamagishi N, Yoshikawa N (2013) A MYB transcription factor controls flower color in soybean. J Hered 104:149–153

Tran LS, Mochida K (2010) Functional genomics of soybean for improvement of productivity in adverse conditions. Funct Integr Genom 10:447–462

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-seq. Bioinformatics 25:1105–1111

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578

Tremblay A, Hosseini P, Li S, Alkharouf NW, Matthews BF (2013) Analysis of *Phakopsora pachyrhizi* transcript abundance in critical pathways at four time-points during infection of a susceptible soybean cultivar using deep sequencing. BMC Genom 14:614

Whitham SA, Eggenberger AL, Zhang C, Chowda-Reddy RV, Martin KM, Hill JH (2015) Recent advances in *in planta* transient expression and silencing systems for soybean using viral vectors. In: Azhakanandam K, Silverstone A, Daniell H, Davey MR (eds) Recent advancements in gene expression and enabling technologies in crop plants. Springer, New York, pp 423–451

Whitham SA, Lincoln LM, Chowda-Reddy RV, Dittman JD, O'Rourke JA, Graham MA (2016a) Virus-induced gene silencing and transient gene expression in soybean (*Glycine max*) using *Bean pod mottle virus* infectious clones. In: Stacey G, Birchler JA, Ecker J, Martin C, Stitt M, Zhou J-M (eds) Current protocols in plant biology. Wiley, Hoboken

Whitham SA, Qi M, Innes RW, Ma W, Lopes-Caitar V, Hewezi T (2016b) Molecular soybean-pathogen interactions. Annu Rev Phytopathol 54:443–468

Wong J, Gao L, Yang Y, Zhai J, Arikit S, Yu Y, Duan S, Chan V, Xiong Q, Yan J, Li S, Liu R, Wang Y, Tang G, Meyers BC, Chen X, Ma W (2014) Roles of small RNAs in soybean defense against *Phytophthora sojae* infection. Plant J 79:928–940

Xie K, Zhang J, Yang Y (2014) Genome-wide prediction of highly specific guide RNA spacers for CRISPR–Cas9-mediated genome editing in model plants and major crops. Mol Plant 7:923–926

Xu X, Zeng L, Tao Y, Vuong T, Wan J, Boerma R, Noe J, Li Z, Finnerty S, Pathan SM, Shannon JG, Nguyen HT (2013) Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. Proc Natl Acad Sci USA 110:13469–13474

Yamada T, Mori Y, Yasue K, Maruyama N, Kitamura K, Abe J (2014) Knockdown of the 7S globulin subunits shifts distribution of nitrogen sources to the residual protein fraction in transgenic soybean seeds. Plant Cell Rep 33:1963–1976

Yamagishi N, Yoshikawa N (2009) Virus-induced gene silencing in soybean seeds and the emergence stage of soybean plants with *Apple latent spherical virus* vectors. Plant Mol Biol 71:15–24

Yamagishi N, Yoshikawa N (2013) Highly efficient virus-induced gene silencing in apple and soybean by *Apple latent spherical virus* vector and biolistic inoculation. Methods Mol Biol 975:167–181

Yang SS, Tu ZJ, Cheung F, Xu WW, Lamb JFS, Jung HJG, Vance CP, Gronwald JW (2011) Using RNA-seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergene cell wall composition in stems. BMC Genom 12:199

Young WP, Schupp JM, Keim P (1999) DNA methylation and AFLP marker distribution in the soybean genome. Theor Appl Genet 99:785–790

Yuan FJ, Zhu DH, Tan YY, Dong DK, Fu XJ, Zhu SL, Li BQ, Shu QY (2012) Identification and characterization of the soybean *IPK1* ortholog of a low phytic acid mutant reveals an exon-excluding splice-site mutation. Theor Appl Genet 125:1413–1423

Zabala G, Vodkin LO (2014) Methylation affects transposition and splicing of a large *CACTA* transposon from a MYB transcription factor regulating anthocyanin synthase genes in soybean seed coats. PLoS ONE 9:e111959

Zabala G, Campos E, Varala KK, Bloomfield S, Jones SI, Win H, Tuteja JH, Calla B, Clough SJ, Hudson M, Vodkin LO (2012) Divergent patterns of endogenous small RNA populations from seed and vegetative tissues of *Glycine max*. BMC Plant Biol 12:177

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829

Zhang C, Ghabrial SA (2006) Development of *Bean pod mottle virus*-based vectors for stable protein expression and sequence-specific virus-induced gene silencing in soybean. Virology 344:401–411

Zhang C, Yang C, Whitham SA, Hill JH (2009) Development and use of an efficient DNA-based viral gene silencing vector for soybean. Mol Plant Microbe Interact 22:123–131

Zhang C, Bradshaw JD, Whitham SA, Hill JH (2010) The development of an efficient multipurpose *Bean pod mottle virus* viral vector set for foreign gene expression and RNA silencing. Plant Physiol 153:52–65

Zhang C, Grosic S, Whitham SA, Hill JH (2012) The requirement of multiple defense genes in soybean *Rsv1*-mediated extreme resistance to *Soybean mosaic virus*. Mol Plant Microbe Interact 25:1307–1313

Zhang C, Whitham SA, Hill JH (2013) Virus-induced gene silencing in soybean and common bean. Methods Mol Biol 975:149–156

Zhong X, Wang Y, Liu X, Gong L, Ma Y, Qi B, Dong Y, Liu B (2009) DNA methylation polymorphism in annual wild soybean (*Glycine soja* Sieb. et Zucc) and cultivated soybean (*G. max* L. Merr.). Can J Plant Sci 89:851–863

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 S-H, 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 T, Schupp JM, Oliphant A, Keim P (1994) Hypomethylated sequences: characterization of the duplicate soybean genome. Mol Genet Genom 244:638–645