Transposon-Based Functional Characterization of Soybean Genes

Devinder Sandhu and Madan K. Bhattacharyya

Abstract
Type II transposable elements that use a cut-and-paste mechanism for jumping from one genomic region to another are ideal for use in tagging and cloning genes. Precise excision from an insertion site in a mutant gene leads to regaining the wild-type function. Thus, the function of a gene can be established based on the mutant phenotype and the regaining of the wild-type phenotype following precise excision of the element. Heterologous type II transposable elements including the Ac/Ds system from maize, the miniature inverted repeat system, mPing from rice, and the Tnt1 retrotransposon from tobacco have been successfully applied in functional analyses of soybean genes. Although several endogenous transposable elements have been identified in soybean, evidence of an active type II transposable element in soybean was largely lacking. We have previously reported the isolation of the type II soybean transposon Tgm9 from intron II of the dihydroflavonol-4-reductase 2 (DFR2) gene of the W4 locus. Tgm9 is an active element and produces variegated flowers through somatic excision. Excision of the Tgm9 element from the progenitor cells of flower buds results in genotypes with purple flowers that are known as germinal revertants. The element was discovered from a commercial soybean cultivar, and the line carrying the element was termed T322. The T322 genome contains only one active Tgm9 copy in the W4 locus. In a recent study, the utility of Tgm9 was assessed by studying a set of random germinal revertants. The new mutations created following excision of Tgm9 from DFR2 were evaluated using a transposon display assay. This
study revealed that Tgm9 transposes to all 20 soybean chromosomes from its original site in the DFR2 gene. Although Tgm9 exhibited preferential transposition to a few genomic regions, across the entire genome 25.7% of the new Tgm9 mutants were detected in exon or intron sequences. Thus, Tgm9 is a suitable endogenous type II transposon to generate an indexed insertional mutant collection for functional characterization of most of the soybean genes.

12.1 Introduction

Soybean is the world’s most valuable crop with high levels of protein (~40%) and oil (~20%), hence it is an important source of human nutrition and livestock and aquaculture feed (Masuda and Goldsmith 2009). It has also become an important source of biodiesel in recent years. In 2014, the USA produced ~35% of the world’s soybean crop with a value approaching $40 billion, and the US export of soybean and soya products including biodiesel was over $30 billion (http://www.soystats.com; http://unitedsoybean.org/article/u-s-soy-exports-hit-new-milestones/). Despite the global economic importance of soybean, the molecular and genetic bases of the physiological processes controlling agronomically important traits are largely unknown. The soybean genome has been sequenced, and expression patterns of soybean genes are mostly known (http://www.soybase.org/SequenceIntro.php; http://soykb.org). However, functional analyses of soybean genes are still very arduous because of the lack of an efficient and rapid transformation procedure.

In the last decade, transposable elements have been successfully utilized in functional characterization of soybean genes. Transposon tagging has been turned out to be an attractive approach in soybean. In this chapter, we focus on the history of transposon tagging approaches in soybean and recent progress in isolating soybean genes using an endogenous transposable element.

12.2 Transposable Elements and Gene Tagging

Transposable elements, originally identified and studied by Barbara McClintock (1956), have become an ideal means of gene isolation. Transposable elements provide unique tools for functional characterization of eukaryotic genomes and have played a major role in understanding gene function and genome organization (Walbot 1992; Martienssen 1998; Peterson 2013; Naito et al. 2014).

After the initial isolation of the bronze locus in maize using the Activator/Dissociation (Ac/Ds) system, many plant genes have been isolated using transposon tagging (Fedoroff et al. 1983; Dooner and Belachew 1989; Jones et al. 1994). The Ac/Ds transposon system was extensively used for gene tagging and functional genomics in maize (Lazarow et al. 2013). However, due to the lack of active and well-characterized transposable elements in many crops, transposon tagging has been limited to a few plant species. Heterologous transposable elements have been exploited in plant species that do not carry any characterized active endogenous transposable element (Bancroft et al. 1993; Chuck et al. 1993; Martienssen 1998). For example, heterologous transposon tagging has been utilized in crop species including tomatoes, tobacco, flax, Arabidopsis, and barley (Jones et al. 1994; Whitham et al. 1994; James et al. 1995; Lawrence et al. 1995; Singh et al. 2012). The most commonly used transposable element
system is Ac/Ds, which transposes to closely linked loci and thus is most suitable for tagging genes in a specific genomic region of a chromosome (Dooner and Belachew 1989).

12.3 Transposition of Ac/Ds in Soybean

The feasibility of using the Ds transposon as an insertional mutagenesis tool was tested in soybean by building a collection of Ds-insertion mutants for different genomic regions through transformation of soybean (Mathieu et al. 2009). As the transformation in soybean is slow, laborious, and expensive, tactical targeting of the transposon is important for maximizing gene-tagging efficiency. To improve the efficiency of functional characterization of soybean genes, an enhancer trap element is added to the T-DNA molecule carrying the Ac/Ds system so that promoter activity of the gene at the insertion site can be monitored (Mathieu et al. 2009). Combining T-DNA with a transposon-based system allowed strategic placement of launch sites in the gene-rich regions. To avoid somatic transposition that can complicate gene cloning, expression of Ac was controlled by a meiosis specific promoter (Mathieu et al. 2009). A total of 900 soybean events were generated, and insertion sites were determined for 200 of them. This revealed that this construct showed a strong bias for insertion into gene-rich regions (Mathieu et al. 2009). This system has been successfully utilized in isolating and characterizing a gene involved in male fertility (Mathieu et al. 2009).

12.4 mPing-Based Mutagenesis in Soybean

mPing is a nonautonomous miniature inverted repeat transposable element (MITE) from rice (Jiang et al. 2003; Kikuchi et al. 2003; Nakazaki et al. 2003). It is a deletion derivative of the Ping element lacking two open reading frames required for transposition (Naito et al. 2009). mPing is an active element that has reached to large copy numbers in some rice lines. For instance, the copy number of mPing is several times higher in Gimbozu compared to its progenitor line Aikoku and related landraces (Naito et al. 2014). In addition to knocking out gene functions through interruption of coding regions, stress-responsive cis-elements of mPing alter expression of adjacent genes (Naito et al. 2009). To study the efficiency of mPing in insertional mutagenesis, mPing and other genes involved in its transposition were transformed into soybean (Hancock et al. 2011). This study showed that mPing produces heritable insertions in soybean over multiple generations. Transposition is developmentally regulated with escalated transposition events observed during late development stages (Hancock et al. 2011). Analysis of the 72 mPing transposition sites in soybean revealed that the insertion sites were distributed on 19 of the 20 soybean chromosomes suggesting that it transposes to unlinked genomic regions (Hancock et al. 2011). Only four insertion sites were located in the annotated pericentromeric region indicating its preference for the euchromatic regions (Hancock et al. 2011). Eighty-five percent of mPing insertions were located within 5 kb, and 51% were within 2.5 kb of an annotated gene (Hancock et al. 2011). In rice, although mPing showed a bias toward gene-rich regions, it was underrepresented in coding sequences, perhaps due to its target site preference for AT-rich regions (Naito et al. 2014). In soybean, mPing did not depict any exon avoidance like in rice, maybe due to lower G/C content in soybean exons as compared to that in rice.

12.5 Tnt1 Retrotransposon Mutagenesis in Soybean

Retrotransposons provide certain advantages over type II transposons and have been used for insertional mutagenesis in several plants (Kumar
The Tnt1 retrotransposon identified from tobacco has been successfully used in plants including Medicago truncatula (d’Erfurth et al. 2003), Arabidopsis (Courtial et al. 2001), lettuce (Mazier et al. 2007), and soybean (Cui et al. 2013) for tagging genes. In soybean, DNA gel blot analysis of Tnt1 in 27 independent transformants revealed that the number of insertions ranged from four to nineteen and were detected in all 20 chromosomes (Cui et al. 2013). Fiber-FISH analysis showed that some of the insertion sites contained single copy Tnt1 insertions, while others contained multiple copy tandem inserts. Tissue culture treatments induced transposition of the element increasing the number of insertions (Cui et al. 2013).

For a successful mutagenesis experiment in a large genome-like soybean, where genes constitute only a small proportion of the total DNA, it is imperative to have transposons that have a preference for gene-rich regions. Analysis of the Tnt1 insertion sites in soybean revealed that it preferentially transposes to gene-containing regions (Cui et al. 2013). Of the 99 Tnt1 sites analyzed, 62% were localized to annotated genes of all 20 chromosomes. As Tnt1 has a preference for gene-containing regions, it is a promising transposon for the functional characterization of the protein-coding genes in soybean (Cui et al. 2013).

Stability and heritability of Tnt1 was tested by growing progenies from plants harboring Tnt1. The locations of Tnt1 among the progenies were stable, and the progenies showed Mendelian inheritance for Tnt1 (Cui et al. 2013). Although Tnt1 transcripts were detected in the vegetative tissues, there was no indication of transposition during development (Cui et al. 2013). Both cotyledonary node and somatic embryogenesis approaches induced activation of Tnt1 among the stable transgenic soybean lines carrying the element (Cui et al. 2013). Thus, through tissue culture of a few transformation events, we can generate a large population of Tnt1-induced mutants.

12.6 Tgm9, An Endogenous Active Transposable Element in Soybean

Application of a well-characterized heterologous transposon in inducing mutation in a target species requires development of a large number of independent transgenic lines with the element. In Ac/Ds system, transposition is mainly to linked regions. This system therefore requires a large collection of transgenic lines with transposons distributed uniformly throughout the genome. Tnt1 requires tissue culture, which may activate endogenous retrotransposons and/or cause epigenomic changes resulting in heritable somaclonal variation. Due to reduced efficiency of transposable elements in the heterologous systems and expenses associated with this approach, the search to identify endogenous active transposons in various plant species continues.

Several endogenous transposable elements have been identified and characterized in soybean (Rhodes and Vodkin 1988; Zabala and Vodkin 2008, 2014; Xu et al. 2010). Of these, Tgm9 is the only known active endogenous transposable element that has been used in gene-tagging experiments. It is a 20,548 bp CACTA-type element that was isolated from the second intron of the dihydrolipoamide-succinyltransferase 2 (DFR2) gene of the W4 locus (Xu et al. 2010). The W4 locus controls pigment formation in flowers and hypocotyls, and the w4-m allele shows variegated flowers due to altered pigment accumulation (Groose et al. 1988; Xu et al. 2010). When the element is excised from DFR2, the petals regain the wild-type purple color. Somatic transposition from petals results in variegated flowers. The w4-m plants produce germinal revertants that carry only purple flowers. In those lines, the excised element transposes into new genetic loci causing heritable mutations; and therefore, the line is called mutable (w4-m). The w4-m line registered as T322 is unstable and undergoes germinal reversion at a high frequency of ~6% per generation (Groose et al. 1990).
12.7 The Mutable Line T322 Contains a Single Active Copy of Tgm9

For a successful transposon tagging experiment, it is important to know the number of active transposon copies in the genome. Two probes, designed from the 5' end and 3' end of Tgm9, were compared with the T322 genome sequence to determine the Tgm9 insertion sites in the T322 genome (Sandhu et al. 2017). A total of 6 and 18 insertion sites were detected by the 5'-end and the 3'-end probes, respectively. Two copies of the element including one in the DFR2 gene on chromosome 17 were detected by both probes. The second copy was found in heterozygous condition and was not detected among mutants previously generated from T322. Therefore, it represents a recent transposition event. Elements detected only by one probe presumably represent truncated copies of Tgm9.

The insertion sites of Tgm9 among some of the mutants were determined by applying a genome walking approach (Sandhu et al. 2017). For st8, a MSFS mutant, Tgm9 insertion site was located in the 10th exon of Glyma.16G072300 that codes for a MER3 DNA helicase involved in crossing over (Baumbach et al. 2016). The Tgm9 insertion site in the ms9 mutant was located in the first intron of Glyma.03G152300 that has no functional annotation (Table 12.1). The Tgm9 insertion sites in Fsp2, Fsp3, and Fsp5 were found in Intron 7 of Glyma.06G174200, Exon2 of Glyma.08G359000, and the 272 bp upstream of Glyma.18G169500, respectively. Glyma.06G174200 encodes a haloacid dehalogenase-like hydrolase, and Glyma.08G359000 and Glyma.18G169500 both encode embryo-specific protein 3, (AT3) (Table 12.1). All three chlorophyll-deficient mutants were due to insertion of Tgm9 in the first intron of Glyma.12G19520 that is predicted to encode a lactate/malate dehydrogenase (Table 12.1).

The genetic linkage map positions of the mutant genes were used to confirm Tgm9 insertion sites in these mutants. The location of Tgm9 insertion sites among the studied mutants agreed with the genetic map positions of the mutant genes; for example, the st8 gene was flanked by two microsatellite markers in a ~62 kb region on Chromosome 16 that contains the MER3 gene tagged by Tgm9 (Raval et al. 2013; Baumbach et al. 2016). The genetic location of ms9 and the physical location of Tgm9 insertion site in the ms9 mutant are on Chromosome 3 (Cervantes-Martinez et al. 2007). Tgm9 in the Fsp2 mutant was found in a gene mapped to a region to which the Fsp2 locus was mapped (Kato and Palmer 2004) (Table 12.1). Similarly, Tgm9 insertion sites in Fsp3 and Fsp5 mutants were localized to Chromosomes 8 and 18, to which the respective mutant genes were previously mapped (Kato and Palmer 2004). The Tgm9 insertion sites in the three chlorophyll-deficient mutants (T323, T325, and T346) were localized to Chromosome 12, ~1.9 Mb from Sat253, to which the three y20 Mdh1-n mutations were mapped (Kato and Palmer 2004).

12.8 Forward Genetics: Identification of Genes by Studying Tgm9-Induced Soybean Mutants

Tgm9 tagging can be used as a tool to conduct forward genetics for isolating soybean genes by using characterized mutant phenotypes generated from insertion of Tgm9. Several mutants such as male-sterile, female-fertile (MSFF) mutant T359 (ms9) (Palmer and Horner 2000) and male-sterile, female-sterile (MSFS) mutants (Kato and Palmer 2003; Palmer et al. 2008a; Raval et al. 2013; Baumbach et al. 2016), partial female-sterile mutants {T364 (Fsp2), T365 (Fsp3), and T367 (Fsp5)} (Kato and Palmer 2004), necrotic root mutants (Palmer et al. 2008b), and chlorophyll-deficient mutants {T323 (y20, Ames 2), T325 (y20, Ames 4), and T346 (y20, Ames 17)} (Palmer et al. 1989) were identified by studying germinal revertants that carry only purple flowers.
<table>
<thead>
<tr>
<th>T #</th>
<th>Mutant phenotype</th>
<th>Gene involved</th>
<th>Upstream/exon/intron/downstream</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T323</td>
<td>y20 (Ames 2)</td>
<td><em>Glyma.J2G159300</em></td>
<td>1st Intron</td>
<td>Lactate/malate dehydrogenase, NAD binding domain</td>
</tr>
<tr>
<td></td>
<td>Mdh1-n (Ames 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T325</td>
<td>y20 (Ames 4)</td>
<td><em>Glyma.J2G159300</em></td>
<td>1st Intron</td>
<td>Lactate/malate dehydrogenase, NAD binding domain</td>
</tr>
<tr>
<td></td>
<td>Mdh 1-n (Ames 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T346</td>
<td>y20 (Ames 17)</td>
<td><em>Glyma.J2G159300</em></td>
<td>1st Intron</td>
<td>Lactate/malate dehydrogenase, NAD binding domain</td>
</tr>
<tr>
<td></td>
<td>Mdh1-n (Ames 19)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T359</td>
<td><em>ms9</em></td>
<td><em>Glyma.03G152300</em></td>
<td>1st intron</td>
<td>No Functional annotation</td>
</tr>
<tr>
<td>T364</td>
<td><em>Fsp2</em></td>
<td><em>Glyma.06G174200</em></td>
<td>7th intron</td>
<td>Haloacid dehalogenase-like hydrolase</td>
</tr>
<tr>
<td>T365</td>
<td><em>Fsp3</em></td>
<td><em>Glyma.08G359000</em></td>
<td>2nd exon</td>
<td>Embryo-specific protein 3, (ATS3)</td>
</tr>
<tr>
<td>T367</td>
<td><em>Fsp5</em></td>
<td><em>Glyma.18G169500</em></td>
<td>272 bp upstream</td>
<td>Embryo-specific protein 3, (ATS3)</td>
</tr>
<tr>
<td>ASR-10-181</td>
<td><em>St8</em></td>
<td><em>Glyma.16G072300</em></td>
<td>10th intron</td>
<td>MER3 DNA helicase</td>
</tr>
</tbody>
</table>

### 12.9 Reverse Genetics: Identification of Tgm9 Insertion Sites Among Germinal Revertants

Following excision of Tgm9 from DFR2, it inserts in new genetic locations resulting in insertion mutations (Xu et al. 2010). We observed that sometimes both copies of Tgm9 excised from DFR2. Therefore, at most two mutations in a single mutant plant can be expected. The element can be used to generate hundreds of thousands of mutations. A high-throughput sequencing of the Tgm9 insertion sites of the population can then identify mutations in a target gene for its functional analyses. This approach is known as reverse genetics.

Seeds of individual T322 plants are harvested, and progenies of each plant are grown in a row. Seeds from a large number of mutable plants with variegated flowers are harvested to identify additional germinal revertants. Progeny rows are grown from individual mutable plants. With the germinal reversion frequency of 6%, we should get about six plants with purple flowers in a progeny row of 100 plants. Plants with only purple flowers are tagged, and leaf tissues are harvested for determining Tgm9 insertion sites (Fig. 12.1).

Recently, we studied Tgm9 insertion sites among 124 germinal revertant plants with purple flowers (Sandhu et al. 2017). Tgm9 transposed multiple times to six genomic regions suggesting its preference for transposition into some genomic regions. About 16.2% of the insertion sites were present in exons, 9.5% in introns, and 2.9%
Fig. 12.1 Experimental plan for obtaining germinal revertants with mutations in novel genetic loci. Approximately 150 progenies of an individual mutable plant are grown in 15-ft-long plots, and a single germinal revertant with only purple flowers tagged for collecting leaf tissues for determining Tgm9 insertion sites and harvesting seeds at maturity.

Fig. 12.2 Percent distribution of Tgm9 insertion sites across the genomes of 105 unique novel soybean mutants. Note that ~28.6% contain insertions in exons, introns or the promoter region and are expected to exhibit complete loss of function (knockout mutations). Other mutations in the genic regions may result in reduced function and are classified as knockdown mutants. Here we consider a genic region to contain a predicted transcript and 2-kb sequences at both 5' and 3'-ends of the transcript. An inter-genic region spans in between two genic regions.

within promoter (300 bp upstream of the transcription initiation site) sequences (Fig. 12.2) (Sandhu et al. 2017). These mutations are most likely knockout mutants suitable for studying gene function.

Tgm9 moves from its original location in the DFR2 locus on Chromosome 17 (Xu et al. 2010) to all 20 soybean chromosomes (Fig. 12.3). The number of insertion sites varied from one (Chromosomes 5, 16, and 19) to sixteen (Chromosome 9) per chromosome (Sandhu et al. 2017). Insertion frequency was higher toward the telomeres. Of the 105 unique insertion sites, 85 (81%) were located in distal 50% of the chromosome arms, and only 20 (19%) were present in proximal half of chromosome arms (Sandhu et al. 2017). Tgm9 showed preference toward euchromatic regions as only 22.9% of the insertion sites were present in the heterochromatic pericentromeric repeat regions, which constitute 52.9% of the soybean genome (Fig. 12.3). Preferential transposition of Tgm9 to the gene-rich regions is particularly important in tagging genes in a large genome-like soybean, where only 16% of the genome contains genes (Sandhu et al. 2017). From the above, it is apparent that the Tgm9 system is suitable for gene identification through forward and reverse genetics studies.

12.10 Transposon Tagging and the Future of Soybean Research

As the soybean genome has been sequenced, the research priority for soybean now is to determine functions of these genes, especially the ones that control the agronomic traits such as stress tolerance, yield, and quality traits. In the last decade, due to addition of a variety of genomic resources in soybean, transposon tagging has emerged as an attractive alternative in determining soybean gene functions. The availability of molecularly characterized transposon-induced soybean mutants will expedite both basic and applied research in this economically important crop. A high-throughput sequencing approach can be used to determine the Tgm9 insertion sites among hundreds of thousands mutants. The
Fig. 12.3 Chromosomal distribution of Tgm9 insertion sites among mutants. Green arrows represent locations of Tgm9 in genic regions, and red arrows represent locations of Tgm9 in inter-genic regions. A purple arrow represents the original location of Tgm9 in DFR2 in the w4-m (T322) line. Centromeres (black rectangles) and heterochromatic regions (gray areas) are shown on individual chromosomes. Scale is represented in million base pairs (Mb) of DNA.

display of those Tgm9 insertion sites in the SoyBase genome browser will facilitate identification of mutants for most single copy soybean genes. Soybean researchers will be able to use this resource to identify suitable mutants for genetically improved soybean for disease and pest resistance, quality and quantity of oil, proteins and seed compositions, and also for adaptation to adverse climatic conditions. Soybean biologists will be able to investigate novel traits unique to soybean as well as conduct translational genomics based on knowledge gained in other model plant species such as Arabidopsis, rice, and maize. The soybean research community will use this resource to gain a better understanding of the molecular basis of physiological processes in soybean. The fundamental knowledge and genetic variation created using Tgm9 will provide means to improve yield and quality in soybean for securing a sustainable supply of this nutritionally important crop in the twenty-first century.

Acknowledgements We are thankful to United Soybean Board and University of Wisconsin at Stevens Point for grant support to conduct the Tgm9 research. We are thankful to David Grant for kindly reviewing the chapter and for his suggestions.

References


