J Biochem Tech (2008) 1(1):11-20
ISSN: 0974-2328

How effective is T-DNA insertional mutagenesis in *Arabidopsis*?

Yi Hong Wang

Received: 2 June 2008 / Received in revised form: 13 June 2008, Accepted: 20 June 2008, Published online: 18 August 2008
©Sevas Educational Society 2008

Abstract

To evaluate the effectiveness of T-DNA insertion in knocking out a gene, 1084 published *Arabidopsis thaliana* insertion mutants representing 755 genes in 648 publications were reviewed. Insertion in the protein-coding region of a gene generates a knockout at least 90% of the time or 25% of the time if it is before the start codon which also produces a knockdown 67% of the time. Insertion after the stop codon had no effect on transcription of the upstream gene 17% of the time compared to 8% in insertion before the start codon. T-DNA insertion can also cause deletion and translocation.

Keywords: *Arabidopsis thaliana*, deletion, expression, knockout, T-DNA insertion, translocation

Introduction

Transfer-DNA (T-DNA) insertion is a highly effective mutagen for genome-wide mutagenesis (Krysan et al. 1999). It has been widely used to produce insertion mutants in *Arabidopsis thaliana* (Alonso et al. 2003; Bechtold et al. 1993; Feldmann 1991; Galbiati et al. 2000; Koncz et al. 1989; Krysan et al. 1999; Rosso et al. 2003; Sessions et al. 2002) for functional characterization of every gene in the genome. Over 360,000 insertions have been mapped in the *Arabidopsis* genome, covering >90% of the genes (Alonso and Ecker, 2006). Because it tends to insert as concatemers (Krysan et al. 1999), most T-DNA insertions result in the loss-of-function alleles although semi-dominant T-DNA mutation has been reported (Bolle et al. 2000) and functional protein is absent in the homozygous mutant plants in most cases examined. Even if mRNA is transcribed, the T-DNA sequence may contains stop codons, resulting in early translation termination (Krysan et al. 1999). Although it is not a perfect technique (Alonso and Ecker 2006; Østergaard and Yanofsky 2004), T-DNA insertion mutagenesis has been a powerful tool to link genes to phenotypes.

Yi Hong Wang

Penn State University
Behrend College School of Science,
4205 College Dr., Erie, PA 16563, USA

Email: yxw23@psu.edu
Phone: 001 814 898 6276
Fax: 001 814 898 6213

Thanks to the availability of these insertion mutants and the complete genome sequence, the number of reports characterizing *Arabidopsis* genes has increased tremendously since 2000. The trend is projected to continue because only a small portion of these mutants have been characterized so far. The purpose of this paper is to summarize the effect of these insertions on gene knockout based on published insertional mutants. To do this, position of insertion site, its effect on transcript/protein level were collected from 648 reports on 755 *Arabidopsis* genes published from 1997 to January 2007 which includes 1084 insertion mutants. This is by no means a complete coverage because not all reports monitored the expression of mutated genes and this may be biased toward insertions that knocked out the target genes. But it includes approximately 90% of all papers characterizing these mutants and may be useful for investigators working with T-DNA mutants.

Effect of T-DNA insertion on expression of mutated genes

The effect of T-DNA insertion is most commonly evaluated by monitoring expression of mutated genes in homozygous insertion mutants. But as described in more detail below, this can be tricky because the transcript level may not be correlated with protein level (Delatte et al. 2005; Pastuglia et al. 2006) which may also depend on position of the insertion in a gene. In some cases, even if the transcript level does not differ significantly from the wild type, still no protein is produced in the mutant (e.g., Monte et al. 2003), complicating the evaluation of T-DNA insertion especially using reverse transcription-polymerase chain reaction (RT-PCR). For RNA gel blot and RT-PCR analyses, probes/primers from downstream and upstream of the insertion site should be used or at least the downstream probe/primers should be used if the insertion is toward the middle of the gene. It is also notable that for all 24 reports examined that employed RT-PCR using primers spanning the insertion site to monitor expression of mutated genes, none

produced any product, presumably because the inserted T-DNA sequence often larger than 5 kb, if transcribed, is too long for normal DNA polymerase to amplify. And if the T-DNA is not transcribed as one unit, it is not usually possible to amplify it by RT-PCR. The exception is that if the insertion is in an intron and the T-DNA has been spliced out (see below). It is also possible that T-DNA inserted into an exon be processed out generating a shorter altered transcript with a very low efficiency (Lehti-Shiu et al. 2005) or only a small part of the T-DNA was inserted (Shin et al. 2004). These are rare cases but RT-PCR using primers spanning the insertion site will yield a product. Overall, insertion in exons or introns can almost always knock out or knock down the target gene (Table 1).

Table 1: Effect of published T-DNA insertion on transcript level of mutated genes in *Arabidopsis*

Insertion site	Number of insertion mutants	No effect on transcript level	Increased transcript level
Exon	609	7 (1.1%)	0 (0%)
Intron	263	2 (0.7%)	0 (0%)
Before start codon	155	17 (11%)	5 (3%)
After stop codon	23	4 (17%)	2 (8%)
Major Deletion	34	0 (0%)	0 (0%)
Total	1084	25 (2.3%)	7 (0.6%)

[a] For each category of insertions, only insertions that showed no or increased effect on expression are listed. The rest are knockout/knockdowns which are not distinguished in exon/intron insertion mutants because a lot of reports used RT-PCR. RT-PCR can show a reduced level of transcript upstream from an insertion site and increased transcript downstream from the insertion site or vice versa, making it difficult to assign the mutant as knockdown or knockup. For "Major Deletions", See Table 5.

Insertions before the start codon or after stop codons are slightly less effective. Because of low correlation between transcript level and protein level as described above, the impact of insertion on expression is more accurately reflected on protein levels, which was used in about 12% (136) of all insertion mutants evaluated. Protein expression of the 136 insertion mutants indicates that the effectiveness for insertions into exon and intron are similar and both are more effective than insertions into either promoter or 5' untranslated regions (5' UTRs---Table 2). When T-DNA is inserted before the start codon, the translation of the gene is most likely reduced rather than completely knocked out (Table 2), hence producing a leaky mutant phenotype. This is illustrated by six insertions in *GPT 1* in which three insertions in the coding region were lethal, while the other three insertions in either 5' or 3' non-coding regions had no effect on the transcript level of *GPT1* (Niewiadomski et al. 2005).

Insertions in exons and introns

As indicated in Tables 1 and 2, insertion into either exon or intron in the protein-coding region is equally effective in knocking out/down the target gene. Among the cases in which insertion failed to change expression pattern of mutated genes, the T-DNA was inserted 39 bp upstream of the stop codon in *BRM* and normal transcription of the gene was not affected (Hurtado et al. 2006). It is not clear, however, why insertion into the exon did not knock out/down the target gene in the other cases. For example, in both *annAt2-1* and *annAt2-2* mutants, the T-DNA was inserted in the last exon of *Arabidopsis annexin 2* (*AnnAt2*) gene. While no transcript was detected by RNA gel blot analysis in *annAt2-1*, the transcript level was similar to that in wild type in *annAt2-2* (Lee et al. 2004). The insertion in *annAt2-2* was more toward the 3' end of the gene but still some distance before the stop codon. It is possible that the transcript detected by Northern in *annAt2-2* is chimeric and just happens to be the similar size as in the case of *ebf2-3* (Gagne et al. 2004). And in two insertion mutants of *AnnAt4*, *annAt4-1* and *annAt4-2*, insertions were also in the similar positions in the last exon and transcript of *AnnAt4* was not detected by RT-PCR in either mutant (Lee et al. 2004).

Table 2: Effect of T-DNA insertion on protein expression of inserted genes in *Arabidopsis*

Insertion site	Exon	Intron	Before start codon	Total
No protein expression	70(88%)	32(82%)	7(41%)	109(80%)
Reduced protein expression	7(8%)	5(13%)	9(53%)	21(16%)
No effect on protein expression	1(1%)	1(3%)	1(6%)	3(2%)
Truncated protein expression	2(2%)	1(3%)	0(0%)	3(2%)
Total	80	39	17	136

It is also puzzling that insertion in the second exon of *CSN5B* failed to alter protein level compared to the wild type although the insertion only changed the phenotype slightly (Dohmann et al. 2005).

If insertion is into an intron, wild type transcript may be produced in such mutants because intron can be spliced out together with the inserted T-DNA sequence. This indeed occurred in insertion mutants of *AGL104* (Verelst et al. 2007), *ASP2* (Miesak and Coruzzi 2002), *AtEXP7* (Cho and Cosgrove 2002), *AtGA2ox6* (Wang et al. 2004), *AtMIS12* (Sato et al. 2005), *ETA2* (Chuang et al. 2004), *FATB* (Bonaventure et al. 2003), *MRH1* (Jones et al. 2006), *SPDS2* (Imai et al. 2004), and *ULT2* (Carles et al. 2005). Although wild type transcripts were produced, it was with decreased efficiency and the level of correctly spliced transcript was reduced. But this reduced level of wild type transcript can be made into protein as demonstrated in *AtMIS12* (Sato et al. 2005) and *ETA2* (Chuang et al. 2004) but not always (e.g., Bonaventure et al. 2003). Such leaky expression of mutated genes due to intron splicing is useful in rescuing an otherwise lethal mutation (Sato et al. 2005). But not all T-DNA insertions into an intron are spliced out (Hurtado et al. 2006) and only a small percentage of insertions in introns produce a reduced level of wild type transcript (10 out of 263 insertions or 4%); the rest are knockouts. Based on protein level, this number is 13% or 5 out of 39 insertions in introns produced reduced levels of wild type protein (Table 2).

If insertion is toward the middle of the gene in an intron or exon, then transcript levels both upstream and downstream of the insertion can be monitored as described earlier. When 32 genes were surveyed for which transcript level was monitored both upstream and downstream of the insertion site, upstream transcript was detected in 29 genes while downstream transcript was detected in only 12 genes. In two cases, neither transripts were detected (Shen et al. 2006; Shimotohno et al. 2006). This is one reason that for insertions in exons/introns, knockout and knockdown are not separated in Table 1. But when truncated transcript is produced due to transcription termination by the insertion (Noh and Amasino

2003; Schnurr et al. 2002), a truncated protein can also be produced (Gusmaroli et al. 2004; Henderson et al. 2005; Kim et al. 2004; Ohtomo et al. 2005; Okushima et al. 2005) which may not be functional (Gusmaroli et al. 2004; Okushima et al. 2005) even though in most cases truncated transcript is not translated (e.g., Dieterle et al. 2005; Doelling et al. 2001; Hashimoto et al. 2005; Schuster et al. 2006; Ullah et al. 2001). Similarly, even if transcripts both up- and down-stream of the insertion site are expressed, still no protein can be produced due to interruption of the coding region by T-DNA (Delatte et al. 2005). In addition, truncated transcript may contain deletions altering protein sequence (see below; Bostick et al. 2004). Finally, it is not clear why one mutant would show no detectable transcript and the others would have truncated transcripts while they were all insertions in the same exon (Noh and Amasino 2003).

Similarly, insertion in the promoter significantly reduced *AGP17* transcript in the roots but not in the leaves compared to the wild type (Gaspar et al 2004), similar to the reduced *MIF1* transcript in the root not the stem when an insertion was 246 bp upstream of the start codon (Hu and Ma 2006). So it is possible that insertion in a promoter could change expression pattern of the downstream gene.

Effectiveness of insertion decreases as it moves further upstream of the 5' end in the promoter/5' UTR based on its effect on transcript level. If insertion before the start codon is grouped according to its distance (in bp) from ATG (Table 3), then knockout rate decreases from 54% to 31% as insertion moves from 1-50 bp toward 201-500 bp upstream of the ATG and knockdown increases at the same time from 33% to 53%.

Table 3: Position effect of T-DNA insertion before the start codon in *Arabidopsis* on transcript level of affected genes[a]

Insertion site	1-50 bp upstream of start codon	51-200 bp upstream of start codon	201-500 bp upstream of start codon	501-1000 bp upstream of start codon	Total
No transcript	21(54%)	20(42%)	10(31%)	9(50%)	60(44%)
Reduced transcript	13(33%)	23(48%)	17(53%)	5(28%)	58(42%)
No effect	3(8%)	2(4%)	4(13%)	3 (17%)	12(9%)
Increased transcript	2(5%)	3(6%)	1(3%)	1 (5%)	7(5%)[b]
Total	39	48	32	18	137

[a] Because 5' UTR and promoter region were not clearly indicated in a lot of characterized insertion mutants, they are listed together. Data gathered from mutants for which an insertion position was given or can be estimated based on information given.
[b]The percentage is higher than that listed in Table 1 because only mutants with insertion position given and less than 1 kb upstream of the start codon were counted here.

Insertion toward the 3' end of a gene but before the stop codon can also effectively knock out the function of the target gene. For example, the insertion at 47 bp upstream of the stop codon in *AHA3* was found to be lethal (Robertson et al. 2004) and a T-DNA insertion at 13 bp before the stop codon also knocked out transcription in *RCI2A* (Mitsuya et al. 2005). On the other hand, as mentioned earlier, insertion at 39 bp before the stop codon in *BRM* did not affect transcript level (Hurtado et al. 2006). Similarly, insertion in the 5' end of a coding region does not necessarily knock out the gene even if it is six bp after start codon. It only led to reduced protein production and less severe phenotype (Pružinská et al. 2007). But overall, this does appear to be the exception rather than the rule. Insertion toward the 3' end of a gene can also produce a weaker phenotype (e.g., Chelysheva et al. 2005; Chen et al 2005; Guo and Ecker 2003; Yoine et al. 2006) although this is not always a case (Robertson et al. 2004).

Insertions before the start codon

Insertion into the promoter region also produces more knockdowns than knockouts. Based on protein expression, among 17 mutants with the T-DNA inserted in the promoter region, 7 showed no (knockout rate 41%) and 9 showed reduced (knockdown rate 53%) expression of mutated genes (Table 2). In the remaining one mutant there was no effect. Compared to the combined knockout (86%) or knockdown (10%) rates in the coding region (both introns and exons) based on protein expression (Table 2), the knockout rate is lower and knockdown rate is higher.

T-DNA insertion into the promoter region can cause misexpression of the downstream gene. Such insertion resulted in fewer *GTE6* transcripts in leaf 7 to a level similar to leaf 4 of the wild-type which caused round laminae in leaf 7, rather than the elliptical laminae of leaf 7 of the wild-type plants. Thus, elevated expression of *GTE6* in leaves 6 and 7 of wild-type plants is important for the development of the elliptical leaf lamina (Chua et al. 2005).

But the trend is reversed when insertion beyond 500 bp is considered, i.e., knockout rate increases to 50% and knockdown rate decreases to 28% (Table 3). Based on this set of data, the number of no-effect insertion also increases (Table 3). This difference in insertion effectiveness explains why 80% of characterized mutants harbor T-DNA insertion in the coding region and only 15% were focused on insertions before the start codon (Table 1). The total knockout rate based on transcript level is 45% and knockdown rate is 41% if all insertions within 1000 bp upstream of the start codon are considered (Table 3). Overall, 11% of insertions before the start codon did not affect transcription of the downstream gene while that number is less than 1% for insertion into the coding region (Table 1). Based on protein level, 6% (1 out of 17) insertions before ATG had no effect on protein expression of the mutant allele (Table 2).

The fact that insertion closer to the start codon is more effective might be because some promoter elements essential for transcription are close (within ~500 bp) to the start codon (Table 4). If T-DNA is inserted into a promoter element such as TATA box (Novillo et al. 2004) or between TATA box and ATG (Nakajima et al. 2004), the gene is likely to be knocked out. The same is true for other elements such as AuxREs (Tatematsu et al. 2004) and G-boxes (Ito et al. 2003). T-DNA insertion between these elements and ATG knocked out or significantly reduced transcription of the respective genes. Even if these elements are farther upstream from the start codon, insertion could still be effective. For example, insertions more than 3 kb upstream of the *UFO* start codon but in CArG box-like sequences that are recognition elements for MADS box DNA-binding proteins disrupted petal development due to reduced *UFO* expression (Durfee et al. 2003). However, insertion of T-DNA among AuxREs in the promoter of the auto regulated *ARF8* increased its transcript level (Goetz et al. 2006).

Insertion upstream of the start codon increased transcription of the downstream gene in several cases. But it is not clear if the increase

is solely driven by the promoter that drives the antibiotic resistance gene in the T-DNA vector (see below). It has been shown that when T-DNA was inserted into the promoter region of *KIS* the fused 35S-Basta-*KIS* fusion transcript abundance increased (Kirik et al. 2002). Similarly, an insertion 100 bp upstream of the start codon in *ARF17* increased its transcript by 7-12-fold compared to wild type, possibly due to the 35S promoter in the T-DNA vector (Sorin et al. 2005). So it is possible that promoter in the T-DNA sequence plays an important role in increasing transcript level of the downstream gene.

downstream gene. A T-DNA was inserted downstream of AT1G65250 but upstream of AT1G65260 (*VIPP1*). Northern blot analysis using both genes as probes detected only significant reduction in *VIPP1* transcription (Kroll et al. 2001). On the other hand, when a T-DNA was inserted 525 bp downstream of *IRT2* but 2440 bp upstream of the start codon of *IRT1*, the expression of *IRT1* was almost completely abolished while that of *IRT2* was also impaired (Varotto et al. 2002). The latter may be caused by T-DNA disruption of the 3' sequence of a gene as discussed above.

Table 4: T-DNA insertion between promoter elements and coding region increases knockout effectiveness in *Arabidopsis*

Gene	*Promoter element*	*Distance to start codon (in bp)*	*T-DNA insertion*	*Effect on Transcript level*	*Reference*
CBF2	TATA box	179	Within TATA box	Expression not detected	Novillo et al. (2004)
SPR1	TATA box	Within 200	Between TATA and ATG	Expression not detected	Nakajima et al. (2004)
IAA19	AuxREs	165	Within AuxREs	Expression not detected	Tatematsu et al. (2004)
APRR9	G-boxes	434-511	Between G-boxes and ATG	Expression severely reduced.	Ito et al. (2003)
Hsp101	Heat-shock elements (HSEs) and TATA box	320-400	Within HSEs but upstream of TATA	No effect on transcription but protein severely reduced	Hong and Vierling (2001)
UFO	CArG box-like sequences	>3 kb	Within CArG box-like sequences	Expression severely reduced.	Durfee et al. (2003)

Insertions after the stop codon

T-DNA insertion after the stop codon is least effective compared to insertion in other parts of a gene. Based on 23 characterized insertions, 17% (or 4) insertions after the stop codon had no effect on transcript level of upstream genes compared to ~1% in insertions in the coding region (Table 1). Its knockout and knockdown rates of 37.5% each are lower than the 45% and 41% for the insertion in before the start codon, respectively. But insertion at some distance downstream from the stop codon still disrupts the transcription of the upstream gene such as *TFL1* (650-700 bp after stop; Ohshima et al. 1997), *CSLA9* (260 bp after stop; Zhu et al. 2003) and *F5H* (283 bp after stop; Ruegger et al. 1999) or interfere with the transcription of the gene such as *LOL1* (629 bp after stop; Epple et al. 2003) and *IRT2* (525 downstream; Varotto et al. 2002). The reason could be that sequence downstream the insertion site contains enhancer or other regulatory sequence essential for expression of the upstream gene. This was demonstrated in *GL1* that harbored a T-DNA insertion 658 bp after the stop codon that separated the coding region from an enhancer element 900 bp downstream of the stop codon, causing a partial phenotype (Larkin et al. 1993; Oppenheimer et al. 1991). Although it is possible that insertion after the stop codons in the above genes were due to separation of downstream regulatory elements by T-DNA, no evidence was reported.

Insertion after stop codons may also knock up the expression of the upstream gene such as *PIP5K9* (Lou et al. 2007) and *ACT7* (Gilliland et al. 2003) or have no apparent effect on the level of transcript such as *ATEM6* (Manfre et al. 2006). Nevertheless, no or reduced level of proteins were detected in these mutants, probably because these mutants also produced shorter than expected transcripts (Gilliland et al. 2003; Manfre et al. 2006). Insertion in the 3' end of a gene could increase its transcript level due to 35S enhancer present in the T-DNA region.

When a T-DNA is inserted in the intergenic region between two genes in the same orientation, it is more likely to disrupt the

T-DNA promoter can drive expression of downstream gene

Most T-DNA sequences contain a selection marker driven by the Cauliflower Mosaic Virus 35S or other promoters which can in turn drive expression of the downstream gene producing a chimeric transcript as mentioned above in *ARF17* (Sorin et al. 2005) and *KIS* (Kirik et al. 2002). This is one factor that is responsible for increased transcript abundance. The 35S promoter was suggested to be responsible for increased *SWP* transcript abundance and size although the mutant phenotype was still recessive, suggesting that the fused T-DNA-*SWP* transcript was not translated or the protein was not functional (Autran et al. 2002). The fused transcript was ~7 kb compared to 5.5 kb in the wild type. Since the insertion was 250 bp upstream of the start codon and at least the first intron was spliced out based on RT-PCR (Autran et al. 2002), it is puzzling as to what factors caused the fused transcript not to be translated other than stop codons. But in *arf19*, the T-DNA was inserted at 12 bp upstream of the start codon and a larger and more abundant mutant transcript was produced which fueled reduced ARF19 synthesis (Willmoth et al. 2005). So transcribed genes driven by the T-DNA promoter may or may not lead to protein synthesis.

Over expression of a mutated gene due to fusion with T-DNA sequence can also occur in insertions in the coding region. In a *swi1* mutant that carried an insertion at 2 bp after the start codon, a T-DNA-*SWI1* fusion transcript was apparently translated with a low efficiency because the mutant phenotype is leaky (Mercier et al. 2001). If T-DNA is inserted into the middle of a gene such as in exon 13 (out of 18) in *AtISA1* as described earlier, the endogenous promoter drives the transcription of sequence upstream the insertion site while promoter in the T-DNA may drive the downstream sequence transcription (Delatte et al. 2005). The downstream transcript was more abundant compared to that of wild type, indicating an over expression due to the T-DNA promoter (see also Bertrand et al. 2005). Despite normal transcription of up- and downstream sequence, no protein was detected because the gene was interrupted by the insertion (Delatte et al. 2005). Over

Table 5: Larger deletions caused by T-DNA insertion in *Arabidopsis* genes.

Gene	*Function*	*Insertion*	*Deletion*	*Reference*
DNA damage repair				
ARS27A	Ribosomal protein S27 /elimination of damaged mRNA	Before ATG	1287 bp including promoter and the 5' UTR	Revenkova et al. 1999
AtLIG4	DNA damage repair	Exon 1	329 bp upstream of insertion including the first 37 codons	Friesner and Britt 2003
AtMSH2	Mismatch repair	Exon 7	1,510 bp from exons 7 to 13 (insertion to the end)	Leonard et al. 2003
AtRAD17	DNA damage repair	Exon 4	172 bp encompassing part of exon 4, intron 4 and a part of exon 5	Heitzeberg et al. 2004
AtRAD51C	DNA damage repair	Intron 2	141 bp of intron 4	Li et al. 2005
DME	Mismatch/ damage repair/ Female Gametophyte Development	Before ATG	177 bp before insertion site	Choi et al. 2002
MRE11	DNA damage repair/ chromosome integrity	Intron 9	The entire exon 10 and portions of introns 9 and 10	Puizina et al. 2004
WEE1	DNA damage response	Exon 1	Most of the coding region (~1.5 kb)	De Schutter et al. 2007
WEE1	DNA damage response	Intron 7	From insertion to the end (~600 bp)	
WEE1	DNA damage response	Exon 9	>330 bp	
Transport				
AtMRP4	Guard cell ABC transporter	Exon 1	−588 to +1545 including a significant portion of the promoter, the entire TMD0, and most of TMD1 domains	Klein et al. 2004
AtNrt2.2	Nitrate transporter	850 after ATG	25 kb from the insertion on including at least one other gene *AtNrt2.1*	Filleur et al. 2001
AtSUC5	Endosperm sucrose transporter	-1061-+1665	-1061 - +1665 including exons 1 and 2	Baud et al. 2005
AtSUC5	Endosperm sucrose transporter	-58-+499	-58 - +499 including most of exon 1	Baud et al. 2005
GORK	Guard cell K channel /transporter	-2280 - +1464	-2280 - +1464 including exon 1 and most of intron 1	Hosy et al. 2003
Cell division				
AGM	Mitosis/Gamete Development	6 bp before ATG	148 bp including 127 bp 5'UTR and 21 bp promoter	Sorensen et al. 2004
ANQ1	MAPKK/ Cytokinesis	Intron 1	5' UTR and exon 1 were deleted and part of *ANQ1* was duplicated	Soyano et al 2003
ASY	Synaptonemal complex assembly	5'UTR	1.1 kb including 74 bp 5' UTR and ~1. 0 kb promoter	Caryl et al. 2000
AtTop6B	Topoisomerase	Exon 12	268 bp consisting of intron 12 to intron 13 (total 18 exons)	Hartung et al. 2002
TUBG2	Spindle formation/ microtubule organization	Exon 2	2440 bp from the insertion to the end of the gene including eight exons	Pastuglia et al. 2006
Light signaling				
APRR5	Circadian rhythm/light sensing	Exon 5	From the insertion to the end and also include At5G24460 downstream	Yamamoto et al. 2003
PHYB	Phytochrome B	Exon 3/Intron 3	~150 bp in the border of exon 3 and intron 3	Reed et al. 1993
PIL5	Phytochrome interacting/seed germination	Exon 2	From insertion to the end of *PIL5*	Oh et al. 2004
Chromatin remodeling				
LHP1	Chromatin remodeling/gene silencing	Up ATG	1.2 kb upstream of the insertion	Gaudin et al 2001
Mom1	SWI2/SNF2 chromatin remodeling	3' end	1,980 bp from the insertion to the end (last 4 exons) to part of promoter of the next gene	Amedeo et al. 2000
Plastid biogenesis				
Alb4	Plastid membrane protein insertion/plastid biogenesis	Intron 6	Exon 7 and intron 7	Gerdes et al. 2006
atTOC34	Plastid protein import/plastid biogenesis	Exon 6	The remaining gene downstream from the insertion (~600 bp)	Constan et al. 2004
Miscellaneous				
AtAPG9	Protein degradation/ autophagy	Intron 3	From insertion to the end covering seven exons	Hanaoka et al. 2002
BAM2	Receptor kinase/male gamete development	Exon 1	>670 bp upstream of the insertion deleted and/or rearranged	DeYoung et al. 2006
NFD1	Mitochondrial ribosomal protein L21/female gametophyte development	Intron 1	138 bp including part of intron 1 and entire exon 2	Portereiko et al. 2006
PRL1	Glucose/ hormone response	Exon 15	~300 bp from inside exon 15 into intron 16	Németh et al. 1998

expression of upstream sequence by endogenous promoter because of the insertion was also reported although the mechanism was not clear (Okushima et al. 2005).

Insertion leads to chimeric transcript

T-DNA insertion can interfere with mRNA processing of affected genes, particularly intron splicing. When the insertion was into an intron within the 5' UTR of *ERS1*, the resulting chimeric transcript contained a segment of T-DNA sequence, a segment of unspliced UTR intron, a segment of native UTR, and the native coding sequence which could not possibly be translated into the functional protein (Wang et al. 2003). Insertion into intron 2 of *WRKY33* produced a larger transcript due to the deletion and duplication at the 3'-end of intron 2 that prevented proper splicing (Zheng et al. 2006). Insertion before the start codon also interferes with correct intron splicing. *PORC* mutant contained an insertion at 26 bp before the start codon and RT-PCR with primers spanning first three exons showed that the introns were not properly spliced (Masuda et al. 2003). In a mutant with a T-DNA insertion in intron 4 (out of 9), the upstream transcript included the first four *ALD1* exons and T-DNA border sequence and that the downstream transcript included a T-DNA sequence at the 5' end and the six 3' exons of *ALD1* at the 3' end (Song et al. 2004). On the other hand, insertion in intron 3 of *SCA3* at +1894 bp produced a transcript with 1667 bp *SCA3* mRNA and more than 270 bp T-DNA left border (Hricova et al. 2006).

Insertion into exons can produce wild type transcript up to the insertion point followed by sequence addition/deletion and T-DNA. If the insertion is in the very last exon, i.e., 47 bp before the stop codon as in *AHA3* (Robertson et al. 2004), the resulting transcript could include the wild type mRNA without the last 65 bases followed by >200 bases of T-DNA. Another insertion in exon 4 (~1/3 of the coding region) yielded a transcript with the first 1/3 wild type mRNA with 17 bases of unknown sequence followed again by >200 bases of T-DNA (Robertson et al. 2004). Longer T-DNA can also be transcribed and attached to the transcript. For example, an insertion at +1787 bp position of *EBF2* had a transcript with 1787 bp *EBF2* mRNA plus ~800 bases of T-DNA (Gagne et al. 2004). In most cases, these chimeric transcripts were not translated or the proteins were not fully functional.

Insertion causes deletion

T-DNA insertion can often cause deletion and rearrangements of the host genome (reviewed by Latham et al. 2006). Out of 1084 insertion mutants surveyed in this paper, about 10% (113) reported deletions although this number can be as high as 87% (Forsbach et al. 2003). But most deletions were small. Among the 113 T-DNA insertion mutants with reported deletions, 82 (73%) had deletions smaller than 100 bp, similar to that reported earlier (78% by Forsbach et al. 2003; 81% by Meza et al. 2002). Of the 82 mutants, 77 (94%) had deletions smaller than 60 bp and 68 (83%) less than 50 bp. With a sample size of 22, Meza et al. (2002) reported that 18 were smaller than 100 bp and 14 of the 18 (78%) were smaller than 60 bp. The average size of deletions is 30 bp among the 82 mutants. The size of deletion for the other 31 mutants ranged from 138 bp to 25 kb but majority was below 2 kb while Forsbach et al. (2003) reported 1 out of 64 (1.6%) and Meza et al. (2002) found 4 out of 22 (18%) deletions are larger than 100 bp.

Interestingly, for the 31 mutants that contained deletions larger than 100 bp (Table 5), 17 (55%) of them were into genes related to DNA functions such as damage repair, replication, chromatin remodeling and meiosis/mitosis and five (16%) into transporter genes. Four of the genes also played roles in male or female gamete development. Similarly, among the 82 mutants with deletions smaller than 100 bp, 29 (39%) insertions were into genes related to DNA repair, replication, meiosis and chromosome remodeling and eight (10%) into transporter genes. These numbers are a lot higher than those expected from random deletions. It is known that repair genes are essential for T-DNA integration which double-stranded DNA breaks can attract (reviewed by Tzfira and Citovsky 2006) and the female reproductive tissues are the primary target of transformation by the *Arabidopsis* floral-dip method (Desfeux et al. 2000). But why are these genes more likely to suffer larger deletions as a result of T-DNA insertion? The answer is not clear although it is conceivable that because T-DNA tends to insert into transcriptionally active regions (Alonso et al. 2003) and these genes are consistently transcribed in the actively-dividing tissue that open themselves up for the T-DNA molecule. The prolonged exposure to T-DNA somehow leads to incorrect double-stranded break which results in larger deletion of the target site. Incidentally, 11 of the 31 mutants had insertions in chromosome V while there was only one in chromosome IV. But the distribution of the other 19 insertions seems to be random among the other three chromosomes.

T-DNA insertion also triggered reciprocal translocation and duplication. Reciprocal translocation was reported for T-DNA insertion into *ARL2* (Guan et al. 2003) and *RHD3* (Yuen et al. 2005) while translocation was observed in *atToc33* (Gutensohn et al. 2004). Insertion in an intron of *WRKY33* caused duplication of the entire gene as described earlier (Zheng et al. 2006). More dramatic rearrangements were reviewed by Latham et al. (2006).

Conclusion

Insertion in the protein-coding region of a gene generates a knockout 86% of the time whether it is in the introns or exons; or 41% of the time if the insertion is before the start codon. Insertion before the start codon produces a knockdown 53% of the time. Insertion after the stop codon had no effect on transcription of the upstream gene 17% of the time compared to 11% in insertion before the start codon. In 10 out of 263 mutants that had an insertion in an intron, the intron was spliced out together with the T-DNA producing wild type transcript at a reduced level. While the absence of the transcript of a gene carrying a T-DNA indicates the gene has been knocked out, it is more complex when an increased transcript level of the gene is detected. Based on this survey, it is more likely than not that protein level is also reduced if it is still synthesized. So it is always advisable to check the transcript level of the mutant allele, especially when a putative homozygous mutant did not show discernible phenotypes. Compared to Northern analysis, RT-PCR can not determine the size of a large transcript created by T-DNA-plant gene fusion produced by insertion upstream of the start codon. The fused transcript can be driven by promoter in the T-DNA and can be chimeric.

The event of T-DNA insertion is known to cause additional changes to the host genome such as deletion and translocation. In these cases, the mutated gene suffers both deletion and insertion, effectively disrupting its expression. Among insertion mutants with reported deletions of the genomic sequence, three quarters of them had deletions smaller than 100 bp. For the large deletions, a disproportionate number (17 out of 31) of them are in genes related to damage repair, replication, chromatin remodeling, meiosis/mitosis, and transport. This may warrant further study to identify possible links between this and mechanism of T-DNA insertion. However, because these conclusions are based on

published T-DNA mutants, it may be biased toward insertions that knock out/down the target genes.

Acknowledgements

The author sincerely thanks Michael A. Campbell, James T. Warren of Penn State University Behrend College and an anonymous reviewer for their valuable suggestions to improve this manuscript. This work was supported by Penn State University Behrend College.

References

Alonso JM, Ecker JR (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. Nat Rev Genet 7:524-536

Alonso JM, Stepanova AN et al. (2003) Genomewide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653 -657

Amedeo P, Habu Y, Afsar K et al (2000) Disruption of the plant gene *MOM* releases transcriptional silencing of methylated genes. Nature 405:203 -206

Autran D, Jonak C, Belcram K et al. (2002) Cell numbers and leaf development in *Arabidopsis*: a functional analysis of the *STRUWWELPETER* gene. EMBO J 21:6036-6049

Baud S, Wuilleme S, Lemoine R et al. (2005) The AtSUC5 sucrose transporter specifically expressed in the endosperm is involved in early seed development in *Arabidopsis*. Plant J 43:824-836

Bechtold N, Ellis J, Pelletier G (1993) *In planta Agrobacterium*-mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. C R Acad Sci Paris 316:1194 -1199

Bertrand C, Benhamed M, Li YF et al. (2005) *Arabidopsis* HAF2 gene encoding TATA-binding protein (TBP)-associated factor TAF1, is required to integrate light signals to regulate gene expression and growth. J Biol Chem 280:1465-1473

Bolle C, Koncz C, Chua N-H (2000) *PAT1*, a new member of the GRAS family, is involved in phytochrome A signal transduction. Genes Dev 14:1269-1278

Bonaventure G, Salas JJ, Pollard MR, Ohlrogge JB (2003) Disruption of the *FATB* gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. Plant Cell 15:1020-1033

Bostick M, Lochhead SR, Honda A, Palmer S, Callis J (2004) Related to ubiquitin 1 and 2 are redundant and essential and regulate vegetative growth, auxin signaling, and ethylene production in *Arabidopsis*. Plant Cell 16:2418-2432

Carles CC, Choffnes-Inada D, Reville K et al. (2005) *ULTRAPETALA1* encodes a SAND domain putative transcriptional regulator that controls shoot and floral meristem activity in *Arabidopsis*. Development 132:897-911

Caryl AP, Armstrong SJ, Jones GH, Franklin FC (2000) A homologue of the yeast *HOP1* gene is inactivated in the *Arabidopsis* meiotic mutant *asy1*. Chromosoma 109:62-71

Chelysheva L, Diallo S, Vezon D et al. (2005) AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis. J Cell Sci 118:4621-4632

Chen C, Zhang W, Timofejeva L, Gerardin Y, Ma H (2005) The *Arabidopsis ROCK-N-ROLLERS* gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation. Plant J 43:321 -334

Cho HT, Cosgrove DJ (2002) Regulation of root hair initiation and expansin gene expression in *Arabidopsis*. Plant Cell 14:3237-3253

Choi Y, Gehring M, Johnson L et al. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell 110:33-42

Chuang HW, Zhang W, Gray WM (2004) *Arabidopsis ETA2*, an apparent ortholog of the human cullin-interacting protein CAND1, is required for auxin responses mediated by the SCF^{TIR1} ubiquitin ligase. Plant Cell 16:1883-1897

Delatte T, Trevisan M, Parker ML, Zeeman SC (2005) *Arabidopsis* mutants *Atisa1* and *Atisa2* have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. Plant J 41:815 -830

Desfeux C, Clough SJ, Bent AF (2000) Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral-dip method. Plant Physiol 123:895-904

De Schutter K, Joubès J, Cools T et al. (2007) *Arabidopsis* WEE1 Kinase Controls Cell Cycle Arrest in Response to Activation of the DNA Integrity Checkpoint. Plant Cell 19:211-225

De Young BJ, Bickle KL, Schrage KJ, Muskett P, Patel K, Clark SE (2006) The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in *Arabidopsis*. Plant J 45:1-16

Dieterle M, Thomann A, Renou JP, Parmentier Y, Cognat V, Lemonnier G, Muller R, Shen WH, Kretsch T, Genschik P (2005) Molecular and functional characterization of *Arabidopsis* Cullin 3A. Plant J 41:386-399

Doelling JH, Yan N, Kurepa J, Walker J Vierstra RD (2001) The ubiquitin-specific protease UBP14 is essential for early embryo development in *Arabidopsis thaliana*. Plant J 27:393-405

Dohmann EMN, Kuhnle C, Schwechheimer C (2005) Loss of the CONSTITUTIVE PHOTOMORPHOGENIC 9 Signalosome Subunit 5 Is Sufficient to Cause the *cop*/*det*/*fus* Mutant Phenotype in *Arabidopsis*. Plant Cell 17:1967-1978

Durfee T, Roe JL, Sessions RA, Inouye C, Serikawa K, Feldmann KA, Weigel D, Zambryski PC (2003) The F-box-containing protein UFO and AGAMOUS participate in antagonistic pathways governing early petal development in *Arabidopsis*. Proc Natl Acad Sci USA 100:8571-8576

Epple P, Mack AA, Morris VR, Dangl JL (2003) Antagonistic control of oxidative stress-induced cell death in *Arabidopsis* by two related, plant-specific zinc finger proteins. Proc Natl Acad Sci USA 100:6831-6836

Feldmann KA (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. Plant J 1:71 -82

Filleur S, Dorbe MF, Cerezo M et al (2001) An *Arabidopsis* T-DNA mutant affected in *Nrt2* genes is impaired in nitrate uptake. FEBS Lett 489:220 -224

Forsbach A, Schubert D, Lechtenberg B, Gils M, Schmidt R (2003) A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. Plant Mol Biol 52:161 -176

Friesner J, Britt AB (2003) Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. Plant J 34:427-440

Gagne JM, Smalle J, Gingerich DJ, et al (2004) *Arabidopsis* EIN3-binding F-box 1 and 2 form ubiquitin-protein ligases that repress ethylene action and promote growth by directing EIN3 degradation. Proc Natl Acad Sci USA 101:6803-6808

Galbiati M, Moreno MA, Nadzan G, Zourelidou M, Dellaporta SL (2000) Large-scale T-DNA mutagenesis in Arabidopsis for functional genomic analysis. Funct Integr Genomics 1:25 -34

Gaspar YM, Nam J, Schultz CJ et al (2004) Characterization of the *Arabidopsis* lysine-rich arabinogalactan-protein *AtAGP17* mutant (*rat1*) that results in a decreased efficiency of *agrobacterium* transformation. Plant Physiol 135:2162-2171

Gaudin V, Libault M, Pouteau S, Juul T, Zhao G, Lefebvre D, Grandjean O (2001) Mutations in *LIKE HETEROCHROMATIN PROTEIN 1* affect flowering time and plant architecture in *Arabidopsis*. Development 128:4847-4858

Gerdes L, Bals T, Klostermann E, Karl M, Philippar K, Hunken M, Soll J, Schunemann D (2006) A second thylakoid membrane-localized Alb3/OxaI/YidC homologue is involved in proper chloroplast biogenesis in *Arabidopsis thaliana*. J Biol Chem 281:16632-16642

Gilliland LU, Pawloski LC, Kandasamy MK, Meagher RB (2003) *Arabidopsis* actin gene *ACT7* plays an essential role in germination and root growth. Plant J 33:319-328

Goetz M, Vivian-Smith A, Johnson SD, Koltunow AM (2006) AUXIN RESPONSE FACTOR8 is a negative regulator of fruit initiation in *Arabidopsis*. Plant Cell 18:1873-1886

Guan C, Rosen ES, Boonsirichai K, Poff KL, Masson PH (2003) The *ARG1-LIKE2* gene of *Arabidopsis* functions in a gravity signal transduction pathway that is genetically distinct from the *PGM* pathway. Plant Physiol 133:100-112

Gusmaroli G, Feng S, Deng XW (2004) The *Arabidopsis* CSN5A and CSN5B subunits are present in distinct COP9 signalosome complexes, and mutations in their JAMM domains exhibit differential dominant negative effects on development. Plant Cell 16:2984-3001

Gutensohn M, Pahnke S, Kolukisaoglu U, Schulz B, Schierhorn A, Voigt A, Hust B, Rollwitz I, Stockel J, Geimer S, Albrecht V, Flugge UI, Klosgen RB (2004) Characterization of a T-DNA insertion mutant for the protein import receptor atToc33 from chloroplasts. Mol Genet Genomics 272:379-396

Guo H, Ecker JR (2003) Plant responses to ethylene gas are mediated by $SCF^{EBF1/EBF2}$-dependent proteolysis of EIN3 transcription factor. Cell 115:667-677

Hanaoka H, Noda T, Shirano Y, Kato T, Hayashi H, Shibata D, Tabata S, Ohsumi Y (2002) Leaf senescence and starvation-induced chlorosis are accelerated by the disruption of an *Arabidopsis* autophagy gene. Plant Physiol 129:1181-1193

Hartung F, Angelis KJ, Meister A, Schubert I, Melzer M, Puchta H (2002) An archaebacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. Curr Biol 12:1787-1791

Hashimoto K, Igarashi H, Mano S, Nishimura M, Shimmen T, Yokota E (2005) Peroxisomal localization of a myosin XI isoform in *Arabidopsis thaliana*. Plant Cell Physiol 46:782-789

Heitzeberg F, Chen IP, Hartung F, Orel N, Angelis KJ, Puchta H (2004) The Rad17 homologue of *Arabidopsis* is involved in the regulation of DNA damage repair and homologous recombination. Plant J 38:954-968

Henderson IR, Liu F, Drea S, Simpson GG, Dean C (2005) An allelic series reveals essential roles for FY in plant development in addition to flowering-time control. Development 132:3597-3607

Hong S-W, Vierling E (2001) Hsp101 is necessary for heat tolerance but dispensable for development and germination in the absence of stress. Plant J 27:25–35

Hosy E, Vavasseur A, Mouline K, Dreyer I, Gaymard F, Poree F, Boucherez J, Lebaudy A, Bouchez D, Very AA, Simonneau T, Thibaud JB, Sentenac H (2003) The *Arabidopsis* outward K^+ channel *GORK* is involved in regulation of stomatal movements and plant transpiration. Proc Natl Acad Sci USA 100:5549-5554

Hricova A, Quesada V, Micol JL (2006) The *SCABRA3* nuclear gene encodes the plastid RpoTp RNA polymerase, which is required for chloroplast biogenesis and mesophyll cell proliferation in *Arabidopsis*. Plant Physiol 141: 942-956

Hu W, Ma H (2006) Characterization of a novel putative zinc finger gene *MIF1*: involvement in multiple hormonal regulation of *Arabidopsis* development. Plant J 45:399–422

Hurtado L, Farrona S, Reyes JC (2006) The putative SWI/SNF complex subunit BRAHMA activates flower homeotic genes in *Arabidopsis thaliana*. Plant Mol Biol 62:291-304

Imai A, Matsuyama T, Hanzawa Y, Akiyama T, Tamaoki M, Saji H, Shirano Y, Kato T, Hayashi H, Shibata D, Tabata S, Komeda Y, Takahashi T (2004) Spermidine synthase genes are essential for survival of *Arabidopsis*. Plant Physiol 135:1565-1573

Ito S, Matsushika A, Yamada H, Sato S, Kato T, Tabata S, Yamashino T, Mizuno T (2003) Characterization of the APRR9 pseudo-response regulator belonging to the APRR1/TOC1 quintet in *Arabidopsis thaliana*. Plant Cell Physiol 44:1237-1245

Jones MA, Raymond MJ, Smirnoff N (2006) Analysis of the root-hair morphogenesis transcriptome reveals the molecular identity of six genes with roles in root-hair development in *Arabidopsis*. Plant J 45:83–100

Kim TH, Kim BH, Yahalom A, Chamovitz DA, von Arnim AG (2004) Translational regulation via 5' mRNA leader sequences revealed by mutational analysis of the *Arabidopsis* translation initiation factor subunit eIF3h. Plant Cell 16:3341-3356

Kirik V, Grini PE, Mathur J, Klinkhammer I, Adler K, Bechtold N, Herzog M, Bonneville JM, Hulskamp M (2002) The *Arabidopsis TUBULIN-FOLDING COFACTOR A* gene is involved in the control of the α/β-tubulin monomer balance. Plant Cell 14:2265-2276

Klein M, Geisler M, Suh SJ, Kolukisaoglu HU, Azevedo L, Plaza S, Curtis MD, Richter A, Weder B, Schulz B, Martinoia E (2004) Disruption of *AtMRP4*, a guard cell plasma membrane ABCC-type ABC transporter, leads to deregulation of stomatal opening and increased drought susceptibility. Plant J 39:219-236

Koncz C, Martini N, Mayerhofer R, Koncz-Kalman Z, Korber H, Redei GP, Schell J (1989) High-frequency T-DNA-mediated gene tagging in plants. Proc Natl Acad Sci USA 86:8467–8471

Kroll D, Meierhoff K, Bechtold N, Kinoshita M, Westphal S, Vothknecht UC, Soll J, Westhoff P (2001) *VIPP1*, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation. Proc Natl Acad Sci USA 98:4238-4242

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an Insertional Mutagen in *Arabidopsis*. Plant Cell 11:2283-2290

Larkin JC, Oppenheimer DG, Pollock S, Marks MD (1993) Arabidopsis *GLABROUS1* gene requires downstream sequences for function. Plant Cell 5:1739-1748

Latham JR, Wilson AK, Steinbrecher RA (2006) The mutational consequences of plant transformation. J Biomed Biotechnol 2006:25376

Lee S, Lee EJ, Yang EJ, Lee JE, Park AR, Song WH, Park OK (2004) Proteomic identification of annexins, calcium-dependent membrane binding proteins that mediate osmotic stress and abscisic acid signal transduction in *Arabidopsis*. Plant Cell 16:1378-1391

Lehti-Shiu MD, Adamczyk BJ, Fernandez DE (2005) Expression of MADS-box genes during the embryonic phase in *Arabidopsis*. Plant Mol Biol 58:89-107

Leonard JM, Bollmann SR, Hays JB (2003) Reduction of stability of *Arabidopsis* genomic and transgenic DNA-repeat sequences (microsatellites) by inactivation of AtMSH2 mismatch-repair function. Plant Physiol 133:328-338

Li W, Yang X, Lin Z, Timofejeva L, Xiao R, Makaroff CA, Ma H (2005) The *AtRAD51C* gene is required for normal meiotic

chromosome synapsis and double-stranded break repair in *Arabidopsis*. Plant Physiol 138:965-976

Lou Y, Gou JY, Xue HW (2007) PIP5K9, an Arabidopsis Phosphatidylinositol Monophosphate Kinase, Interacts with a Cytosolic Invertase to Negatively Regulate Sugar-Mediated Root Growth. Plant Cell 19:163-181

Manfre AJ, Lanni LM, Marcotte WR Jr (2006) The *Arabidopsis* group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development. Plant Physiol 140:140-149

Masuda T, Fusada N, Oosawa N et al (2003) Functional analysis of isoforms of NADPH: protochlorophyllide oxidoreductase (POR), PORB and PORC, in *Arabidopsis thaliana*. Plant Cell Physiol 44:963-974

Mercier R, Vezon D, Bullier E, Motamayor JC, Sellier A, Lefèvre F, Pelletier G, Horlow H (2001) SWITCH1 (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis. Genes Dev 14:1859-1871

Meza TJ, Stangeland B, Mercy IS, Skarn M, Nymoen DA, Berg A, Butenko MA, Hakelien AM, Haslekas C, Meza-Zepeda LA, Aalen RB (2002) Analyses of single-copy *Arabidopsis* T-DNA-transformed lines show that the presence of vector backbone sequences, short inverted repeats and DNA methylation is not sufficient or necessary for the induction of transgene silencing.. Nucleic Acids Res 30:4556-4566

Miesak BH, Coruzzi GM (2002) Molecular and physiological analysis of *Arabidopsis* mutants defective in cytosolic or chloroplastic aspartate aminotransferase. Plant Physiol 129:650-660

Mitsuya S, Taniguchi M, Miyake H, Takabe T (2005) Disruption of *RCI2A* leads to over-accumulation of Na^+ and increased salt sensitivity in *Arabidopsis thaliana* plants. Planta 222:1001-1009

Monte E, Alonso JM, Ecker JR, Zhang Y, Li X, Young J, Austin-Phillips S, Quail PH (2003) Isolation and characterization of *phyC* mutants in *Arabidopsis* reveals complex crosstalk between phytochrome signaling pathways. Plant Cell 15:1962-1980

Nakajima K, Furutani I, Tachimoto H, Matsubara H, Hashimoto T (2004) *SPIRAL1* encodes a plant-specific microtubule-localized protein required for directional control of rapidly expanding *Arabidopsis* cells. Plant Cell 16:1178-1190

Németh K, Salchert K, Putnoky P, Bhalerao R, Koncz-Kálmán Z, Stankovic-Stangeland B, Bakó L, Mathur J, Ökrész L, Stabel S, Geigenberger P, Stitt M, Rédei GP, Schell J, Koncz C (1998) Pleiotropic control of glucose and hormone responses by PRL1, a nuclear WD protein, in *Arabidopsis*. Genes Dev 12:3059-3073

Niewiadomski P, Knappe S, Geimer S et al (2005) The *Arabidopsis* plastidic glucose 6-phosphate/phosphate translocator GPT1 is essential for pollen maturation and embryo sac development. Plant Cell 17:760-775

Noh YS, Amasino RM (2003) *PIE1*, an ISWI family gene, is required for *FLC* activation and floral repression in *Arabidopsis*. Plant Cell 15:1671-1682

Novillo F, Alonso JM, Ecker JR, Salinas J (2004) CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in *Arabidopsis*. Proc Natl Acad Sci USA 101:3985-3990

Oh E, Kim J, Park E, Kim JI, Kang C, Choi G (2004) PIL5, a phytochrome-interacting basic helix-loop-helix protein, is a key negative regulator of seed germination in *Arabidopsis thaliana*. Plant Cell 16:3045-3058

Ohshima S, Murata M, Sakamoto W, Ogura Y, Motoyoshi F (1997) Cloning and molecular analysis of the *Arabidopsis* gene *Terminal Flower 1*. Mol Gen Genet 254:186-194

Ohtomo I, Ueda H, Shimada T, Nishiyama C, Komoto Y, Hara-Nishimura I, Takahashi T (2005) Identification of an allele of *VAM3/SYP22* that confers a semi-dwarf phenotype in *Arabidopsis thaliana*. Plant Cell Physiol 46:1358-1365

Okushima Y, Mitina I, Quach HL, Theologis A (2005) AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. Plant J 43:29–46

Oppenheimer DG, Herman PL, Sivakumaran S, Esch J, Marks MD (1991) A *myb* gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. Cell 67:483-493

Østergaard L, Yanofsky MF (2004) Establishing gene function by mutagenesis in *Arabidopsis thaliana*. Plant J 39:682-696

Pastuglia M, Azimzadeh J, Goussot M, Camilleri C, Belcram K, Evrard JL, Schmit AC, Guerche P, Bouchez D (2006) γ-tubulin is essential for microtubule organization and development in *Arabidopsis*. Plant Cell 18:1412-1425

Portereiko MF, Sandaklie-Nikolova L, Lloyd A, Dever CA, Otsuga D, Drews GN (2006) *NUCLEAR FUSION DEFECTIVE1* encodes the *Arabidopsis* RPL21M protein and is required for karyogamy during female gametophyte development and fertilization. Plant Physiol 141:957-965

Pružinská A, Anders I, Aubry S, Schenk N, Tapernoux-Lüthi E, Müller T, Kräutler B, Hörtensteiner S (2007) In vivo participation of red chlorophyll catabolite reductase in chlorophyll breakdown. Plant Cell 19:369-387

Puizina J, Siroky J, Mokros P, Schweizer D, Riha K (2004) Mre11 deficiency in *Arabidopsis* is associated with chromosomal instability in somatic cells and Spo11-dependent genome fragmentation during meiosis. Plant Cell 16:1968-1978

Reed JW, Nagpal P, Poole DS, Furuya M, Chory J (1993) Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development. Plant Cell 5:147-157

Revenkova E, Masson J, Koncz C, Afsar K, Jakovleva L, Paszkowski J (1999) Involvement of *Arabidopsis* thaliana ribosomal protein S27 in mRNA degradation triggered by genotoxic stress. EMBO J 18:490-499

Robertson WR, Clark K, Young JC, Sussman MR (2004) An *Arabidopsis thaliana* Plasma Membrane Proton Pump Is Essential for Pollen Development. Genetics 168:1677-1687

Rosso MG, Li Y, Strizhov N, Reis B, Dekker K, Weisshaar B (2003) An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol 53:247–259

Ruegger M, Meyer K, Cusumano JC, Chapple C (1999) Regulation of ferulate-5-hydroxylase expression in *Arabidopsis* in the context of sinapate ester biosynthesis. Plant Physiol 119:101-110

Sato H, Shibata F, Murata M (2005) Characterization of a Mis12 homologue in *Arabidopsis thaliana*. Chromosome Res 13:827–834

Schnurr JA, Shockey JM, de Boer GJ, Browse JA (2002) Fatty acid export from the chloroplast. Molecular characterization of a major plastidial acyl-coenzyme A synthetase from *Arabidopsis*. Plant Physiol 129:1700-1709

Schuster J, Knill T, Reichelt M, Gershenzon J, Binder S (2006) Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. Plant Cell 18:2664-2679

Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, Clarke JD, Cotton D, Bullis D, Snell J, Miguel T, Hutchison D, Kimmerly B, Mitzel T, Katagiri F, Glazebrook J, Law M, Goff SA. (2002) A high-throughput *Arabidopsis* reverse genetics system. Plant Cell 14:2985–2994

Shen W, Wei Y, Dauk M, Tan Y, Taylor DC, Selvaraj G, Zou J (2006) Involvement of a glycerol-3-phosphate dehydrogenase in modulating the NADH/NAD^+ ratio provides evidence of a mitochondrial glycerol-3-phosphate shuttle in *Arabidopsis*. Plant Cell 18:422-441

Shimotohno A, Ohno R, Bisova K, Sakaguchi N, Huang J, Koncz C, Uchimiya H, Umeda M (2006) Diverse phosphoregulatory mechanisms controlling cyclin-dependent kinase-activating kinases in *Arabidopsis*. Plant J 47:701–710

Shin H, Shin HS, Dewbre GR, Harrison MJ (2004) Phosphate transport in *Arabidopsis*: Pht1;1 and Pht1;4 play a major role in phosphate acquisition from both low- and high-phosphate environments. Plant J 39:629-642

Song JT, Lu H, Greenberg JT (2004) Divergent roles in *Arabidopsis thaliana* development and defense of two homologous genes, *ABERRANT GROWTH AND DEATH2* and *AGD2-LIKE DEFENSE RESPONSE PROTEIN1*, encoding novel aminotransferases. Plant Cell 16:353-366

Sorensen AM, Kroeber S, Saedler H (2004) The *ABNORMAL GAMETOPHYTES* (*AGM*) gene product of *Arabidopsis* demonstrates a role in mitosis during gamete development. Plant Cell Physiol 45:905-913

Sorin C, Bussell JD, Camus I, Ljung K, Kowalczyk M, Geiss G, McKhann H, Garcion C, Vaucheret H, Sandberg G, Bellini C (2005) Auxin and light control of adventitious rooting in *Arabidopsis* require ARGONAUTE1. Plant Cell 17:1343-1359

Soyano T, Nishihama R, Morikiyo K, Ishikawa M, Machida Y (2003) NQK1/NtMEK1 is a MAPKK that acts in the NPK1 MAPKKK-mediated MAPK cascade and is required for plant cytokinesis. Genes Dev 17:1055-1067

Tatematsu K, Kumagai S, Muto H, Sato A, Watahiki MK, Harper RM, Liscum E, Yamamoto KT (2004) *MASSUGU2* encodes Aux/IAA19, an auxin-regulated protein that functions together with the transcriptional activator NPH4/ARF7 to regulate differential growth responses of hypocotyl and formation of lateral roots in *Arabidopsis thaliana*. Plant Cell 16:379-393

Tzfira T, Citovsky V (2006) *Agrobacterium*-mediated genetic transformation of plants: biology and biotechnology. Curr Opin Biotechnol 17:147-154

Ullah H, Chen JG, Young JC, Im KH, Sussman MR, Jones AM (2001) Modulation of cell proliferation by heterotrimeric G protein in *Arabidopsis*. Science 292:2066-2069

Varotto C, Maiwald D, Pesaresi P, Jahns P, Salamini F, Leister D (2002) The metal ion transporter IRT1 is necessary for iron homeostasis and efficient photosynthesis in *Arabidopsis thaliana*. Plant J 31:589–599

Verelst W, Saedler H, Munster T (2007) MIKC* MADS-Protein Complexes Bind Motifs Enriched in the Proximal Region of Late Pollen-Specific *Arabidopsis* Promoters. Plant Physiol 143:447-460

Wang H, Caruso LV, Downie AB, Perry SE (2004) The embryo MADS domain protein AGAMOUS-Like 15 directly regulates expression of a gene encoding an enzyme involved in gibberellin metabolism. Plant Cell 16:1206-1219

Wang W, Hall AE, O'Malley R, Bleecker AB (2003) Canonical histidine kinase activity of the transmitter domain of the ETR1 ethylene receptor from *Arabidopsis* is not required for signal transmission. Proc Natl Acad Sci USA 100:352-357

Wilmoth JC, Wang S, Tiwari SB, Joshi AD, Hagen G, Guilfoyle TJ, Alonso JM, Ecker JR, Reed JW (2005) NPH4/ARF7 and ARF19 promote leaf expansion and auxin-induced lateral root formation. Plant J 43:118–130

Yamamoto Y, Sato E, Shimizu T, Nakamich N, Sato S, Kato T, Tabata S, Nagatani A, Yamashino T, Mizuno T (2003) Comparative genetic studies on the *APRR5* and *APRR7* genes belonging to the APRR1/TOC1 quintet implicated in circadian rhythm, control of flowering time, and early photomorphogenesis. Plant Cell Physiol 44:1119-1130

Yoine M, Nishii T, Nakamura K (2006) *Arabidopsis* UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. Plant Cell Physiol 47:572-580

Yuen CY, Sedbrook JC, Perrin RM, Carroll KL, Masson PH (2005) Loss-of-function mutations of *ROOT HAIR DEFECTIVE3* suppress root waving, skewing, and epidermal cell file rotation in *Arabidopsis*. Plant Physiol 138:701-714

Zhang X, Myers AM, James MG (2005) Mutations affecting starch synthase III in *Arabidopsis* alter leaf starch structure and increase the rate of starch synthesis. Plant Physiol 138:663-674

Zheng Z, Qamar SA, Chen Z, Mengiste T (2006) *Arabidopsis* WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. Plant J 48:592-605

Zhu Y, Nam J, Carpita NC, Matthysse AG, Gelvin SB (2003) *Agrobacterium*-mediated root transformation is inhibited by mutation of an *Arabidopsis* cellulose synthase-like gene. Plant Physiol 133:1000-1010