Plant Cell Physiol. 46(1): 14–22 (2005)
doi:10.1093/pcp/pci502, available online at www.pcp.oupjournals.org

Mini Review

Molecular Genetics Using T-DNA in Rice

Gynheung An [1,3], Shinyoung Lee [1], Sung-Hyun Kim [2] and Seong-Ryong Kim [2]

[1] *National Research Laboratory of Plant Functional Genomics, Division of Molecular and Life Sciences, Pohang University of Science and Technology (POSTECH), Pohang 790-784, Korea*
[2] *Department of Life Science, Sogang University, Seoul 121-742, Korea*

Now that sequencing of the rice genome is nearly completed, functional analysis of its large number of genes is the next challenge. Because rice is easy to transform, T-DNA has been used successfully to generate insertional mutant lines. Collectively, several laboratories throughout the world have established at least 200,000 T-DNA insertional lines. Some of those carry the *GUS* or *GFP* reporters for either gene or enhancer traps. Others are activation tagging lines for gain-of-function mutagenesis when T-DNA is inserted in the intergenic region. A forward genetic approach showed limited success because of somaclonal variations induced during tissue culture. To utilize these resources more efficiently, tagged lines have been produced for reverse genetics approaches. DNA pools of the T-DNA-tagged lines have been prepared for polymerase chain reaction (PCR) screening of insertional mutants in a given gene. Appropriate T-DNA insertion sites are determined by sequencing the region flanking the T-DNA. This information is then used to make databases that are shared with the scientific community. International efforts on seed amplification and maintenance are needed to exploit these valuable materials efficiently.

Keywords: DNA pools — Functional genomics — Gene tagging — Reverse genetics — Tag end sequence — T-DNA.

Abbreviations: CaMV, cauliflower mosaic virus; GFP, green fluorescent protein; GUS, β-glucuronidase; TAIL PCR, thermal asymmetric interlaced polymerase chain reaction; T-DNA, transferred DNA; TES, tag end sequence.

Introduction

More than 40,000 putative genes have been identified from sequencing of the rice genome (Feng et al. 2002, Goff et al. 2002, Sasaki et al. 2002, Yu et al. 2002). Therefore, the most challenging goal now is to discover their functional roles. To facilitate such an evaluation, several approaches have been developed. In a broad sense, the first reverse genetics study was reported in 1988 when an antisense approach was taken to introduce a ribulose bisphosphate carboxylase small subunit gene into a plant system (Rodermel et al. 1988). Current methods include antisense and RNA interference (RNAi) technologies (Chuang and Meyerowitz 2000) to decrease the expression of target genes, as well as insertional mutagenesis to knock out gene expression (Feldmann 1991, Jeon et al. 2000). Among these, random insertional mutagenesis by T-DNA or transposons has been most widely used for large-scale analyses. This technique is not only efficient for identifying knockout mutants, but can also be employed for both promoter and activation tagging. The recent establishment of a large number of insertional mutants will accelerate these reverse genetics approaches for studying rice gene function in this model monocot species. In addition, homologous recombination (reviewed by Hanin and Paszkowski 2003) and viral-induced gene silencing (reviewed by Robertson 2004) are apparently effective techniques. The former has been applied in rice; however, the efficiency is low and examples are limited. The latter has not been applied successfully in rice until now. Tilling (targeting induced local lesions in genomes) is an obvious, attractive reverse genetics strategy that can be adapted in rice for finding point mutations in genes of interest (reviewed by Henikoff et al. 2004). Our focus in this review is on reverse genetics using T-DNA in rice.

Generation of Knockout Mutants by T-DNA

T-DNA has been broadly utilized for generating insertional mutant pools in *Arabidopsis* (Feldmann 1991, Azpiroz-Leehan and Feldmann 1997, Bouche and Bouchez 2001). For example, the *Arabidopsis* knockout facility at the University of Wisconsin has established a population of 60,480 T-DNA-tagged lines (Krysan et al. 1999). Sessions et al. (2002) have reported the generation of 100,000 T-DNA-transformed lines, and Szabados et al. (2002) have analyzed the distribution of 1,000 T-DNA sequence tags isolated from their T-DNA insertion lines. Recently, >225,000 independent T-DNA insertional lines of *Arabidopsis* were created that represent almost the entire gene space (Alonso et al. 2003). Compared with transposons, T-DNA insertions appear to be relatively randomly distributed in a plant's genome (Kolesnik et al. 2004, Sallaud et al.

---

[3] Corresponding author: E-mail, genean@postech.ac.kr; Fax, +82-54-279-0659.

Table 1 Inventory of T-DNA insertional mutants in rice

Character	No. of lines	Web site	Reference
Gene trap	50,000	http://www.postech.ac.kr/life/pfg	Jeon et al., 2000
Activation/gene trap	50,000	http://www.postech.ac.kr/life/pfg	Jeong et al. 2002
Enhancer trap	46,000	http://genoplante-info.infobiogen.fr/oryzatagline/project.htm	Sallaud et al. 2003, Sallaud et al. 2004
Enhancer trap	42,000	http://www.ricefgchina.org	Wu et al. 2003

2004). Therefore, this mutagen is suitable for achieving near-saturation mutagenesis. However, it is prerequisite that the genome size is not too large and that transformation frequency is high enough to ensure the generation of a large number of transgenic plants. Currently, *Arabidopsis* and rice are the only plant species that meet such demands.

In *Arabidopsis*, the *in planta* transformation method is used to generate insertional mutant populations. Therefore, obtaining a large number of transformants has been relatively easy. However, that technique has not been applied routinely to other species. In rice, the *Agrobacterium*-mediated co-cultivation method is efficient enough to produce a large quantity of T-DNA insertional mutant plants (Hiei et al. 1994, Lee et al. 1999, Sallaud et al. 2003). For example, the research group of An has generated approximately 100,000 fertile rice lines tagged by T-DNA (Jeon et al. 2000, Jeong et al. 2002, Ryu et al. 2004). In addition, work by Zhang and colleagues has resulted in >30,000 T-DNA insertional lines (Wu et al. 2003). Several other groups independently have produced T-DNA insertional mutant lines (Yin and Wang 2000, Chen et al. 2003, Sallaud et al. 2003, Sallaud et al. 2004, Sha et al. 2004) so that, altogether, >200,000 T-DNA tags have now been generated in this species. Representative T-DNA tagging populations are presented in Table 1. These populations are large enough to find a knockout in a given gene at >90% probability, assuming that T-DNA is randomly inserted into a chromosome. However, the number of tagged lines necessary for saturating all the rice genes may actually be smaller than the estimated value because T-DNA insertions occur preferentially in gene-rich regions (Barakat et al. 2000, An et al. 2003, Chen et al. 2003, Wu et al. 2003, Sallaud et al. 2004).

Generation of Activation Tagging Mutants

About two-thirds of the *Arabidopsis* genome is duplicated in the form of large chromosomal segments (Arabidopsis Genome Initiative 2000). In addition, about 4,000 genes are tandemly repeated as two or more copies. For rice, the fraction of locally duplicated genes ranges from 15.4 to 30.4%, depending on the chromosome (Goff et al. 2002). When the sequences showing a minimum of 80% identity over 100 bp were considered to be duplicated sequences, 851 among 2,000 rice cDNA markers (41%) are single loci, 509 (24%) have two copies and the remainder (35%) have three or more (Goff et al. 2002). The greatest duplication has been observed on chromosomes 11 and 12 (Nagamura et al. 1995, Harushima et al. 1998, Wilson et al. 1999). Therefore, unless multiple gene-mutated plants are generated, T-DNA tagging or transposon tagging proves inefficient in studying these duplicated genes. It is estimated that fewer than 10% of the genes tagged in *Arabidopsis* and rice are likely to generate a visible phenotypic change (Feldmann 1991, Bouche and Bouchez 2001, Jeon and An 2001).

The activation tagging approach is one method that complements the technologies needed for studying genes whose function cannot be resolved by insertional mutagenesis because of gene redundancy, a lethal phenotype in loss-of-function mutants, or phenotype expression only under specific conditions. In addition, activation tagging generates gain-of-function mutants that have several advantages over knockout mutants. First, it is dominant, making it possible to analyze the function of duplicated genes. Secondly, its mutant spectrum differs from that of loss-of-function mutants, often generating beneficial traits for crop improvement. An alternative to activation tagging is the use of tissue-specific promoters that ectopically express or increase the expression of target genes only during certain life cycle phases or in targeted tissues (Weigel et al. 2000). For example, a chemically inducible activation tagging system has been used recently to identify genes controlling the vegetative-to-embryonic transition in *Arabidopsis* (Zuo et al. 2002). A heat shock tagging system has also been developed in *Arabidopsis* to induce transcription of flanking genomic sequences in response to heat treatment (Matsuhara et al. 2000).

Activation tagging, using T-DNA inserts, has been applied effectively in *Arabidopsis* (Kardailsky et al. 1999, Ito and Meyerowitz 2000, Lee et al. 2000, Weigel et al. 2000, Kirik et al. 2004, Tani et al. 2004). T-DNA carries strong activator elements that enhance the expression of genes adjacent to the insertion site. The cauliflower mosaic virus (CaMV) 35S enhancer or promoter sequences (Odell et al. 1985) have been implemented most frequently as a system activator. Their primary function is to overexpress tagged genes to reveal dominant gain-of-function phenotypes. It is unclear whether the enhancer acts by quantitatively increasing the original expression patterns or by promoting ectopic or constitutive overexpression of the nearby gene (Neff et al. 1999, van der Graaff et al. 2000). From the T-DNA activation tagging pools of *Arabidopsis*, Weigel et al. (2000) have characterized >30 dominant mutants with various phenotypes and have shown that in several of the mutants the tagging vector causes overexpression of the gene immediately adjacent to the inserted enhancer. Finally, activation tagging has been employed successfully not only to study plant growth and development, but also to discover disease resistance genes in *Arabidopsis* (reviewed by Tani et al. 2004).

Jeong and colleagues have demonstrated that the CaMV 35S enhancer element efficiently increases expression of nearby genes in rice. They have developed a binary vector that carries the tetramerized 35S enhancers within T-DNA. This vector has been used to generate activation tagging pools for >50,000 individual transformants (Jeong et al. 2002). Examination of randomly selected tags in the intergenic regions has shown that approximately 40% of the enhancer insertions increase nearby gene expression. However, enhanced expression does not always result in phenotypic alteration (Neff et al. 1999, Weigel et al. 2000, Jeong et al. 2002). This differs from ectopic expression by strong promoters, e.g. those for actin or ubiquitin, which cause constitutive overexpression in all tissues (McElroy et al. 1990, Cornejo et al. 1993, LeClere and Bartel 2001). Nonetheless, the dominant mutation frequency of an activation tagging population is much higher than that of a tag population generated by a simple insertion vector, thereby indicating that the activation tagging strategy works in rice (Jeong et al. 2002). In fact, characterization of these dominant mutants has led to the identification of a number of genes tagged by the enhancer in rice (Jeong et al. unpublished data).

Entrapment Tagging

Entrapment systems can be divided into three groups: enhancer trap; promoter trap; and gene trap (Springer 2000). In these systems, T-DNA can be engineered to carry a reporter gene next to the T-DNA border (Jefferson et al. 1987, Pang et al. 1996, Jeon et al. 2000, Ryu et al. 2004). In the enhancer trap system, the reporter gene is fused to a minimal promoter, which is activated when T-DNA is inserted near an enhancer element (Klimyuk et al. 1995, Campisi et al. 1999). Enhancer trapping yields a higher frequency of in situ reporter gene activation. In the promoter trap system, a promoter-less reporter gene is placed next to the T-DNA border. When the reporter is inserted into an exon, the knockout gene and reporter form a translational fusion (Topping and Lindsey 1997). In contrast, gene trap constructs possess an intron with multiple splice donor and acceptor sequences in front of the reporter gene (Sundaresan et al. 1995). This allows splicing from the donor sites in the disrupted gene to the acceptor sites in the reporter gene, resulting in fusion of the upstream exon sequences to the reporter regardless of insert position. Therefore, expression patterns of the tagged genes can be monitored by assaying reporter gene activity.

The gene for β-glucuronidase (*GUS*) has frequently served as a reporter because of the easy detection of its gene product and the tolerance for N-terminal translational fusions in its enzyme activity (Jefferson et al. 1987). Although this reporter has been used to generate T-DNA tagging populations in rice (Jeon et al. 2000), there are some drawbacks to this method. Samples are destroyed during the assay staining and de-staining procedures. In addition, availability of the substrate is not equal for all cell types, resulting in false data. Therefore, non-invasive and non-destructive reporter genes, such as the genes for green fluorescent protein (*GFP*) or luciferase, have been used for entrapment in plants. Tagging populations with the *GFP* gene have now been established for *Arabidopsis* and rice (Haseloff et al. 1997, Ryu et al. 2004). With its rapid turnover and half-life of 3 h, luciferase is well suited as a real-time reporter for *in planta* gene expression studies (Thompson et al. 1991, Millar et al. 1992). Recently, enhancer trap systems in tomato and *Arabidopsis* have demonstrated the usefulness of the reporter in random gene tagging experiments (Meissner et al. 2000, Alvarado et al. 2004). However, one of the weaknesses of these reporter systems using the genes for GFP or luciferase is their sensitivity, because activity is not cumulative. Therefore, genes with weak expression are not detected when using the non-destructive reporters. In addition, the luciferase assay requires a substrate, making it difficult to conduct in situ assays.

As many as 30% of the tagged lines show activation of the reporter gene in the *Arabidopsis* promoter trap system (Sundaresan 1996). In the rice gene trap system, at least 10% of the tagged lines are GUS positive in the roots, leaves, flowers or seeds (Jeon et al. 2000). Some tags are tissue or organ specific, while others are ubiquitous in all examined organs. If one includes reporter gene activation caused by certain environmental conditions or chemicals, such as growth substances, total tagging efficiency is higher. Chin et al. (1999) reported that GUS expression was detected in 8% of the *Ds*-carrying lines in various parts of rice panicles.

An enhancer trap system has been developed for rice (Greco et al. 2003, Wu et al. 2003, Ito et al. 2004). Specifically, the modified enhancer system uses the yeast transcription factor GAL4 (Wu et al. 2003). This system has been applied already not only to study development in *Drosophila* (Brand and Perrimon 1993), but also to rescue mutations or to express a toxin in specific cells and tissues (Brand and Perrimon 1993, Phelps and Brand 1998). There, the activator domain of GAL4 is replaced by the activation domain of herpes simplex virus protein 16 (Triezenberg et al. 1988). This GAL4/VP16 system has been applied successfully during the construction of enhancer trap lines in rice (Wu et al. 2003). In this study, GUS expression in 25–59% of the enhancer trap transformants has been detected in the various tissues. Furthermore, enhancer trap lines have been generated using the *Ac/Ds* system and about 10% of the tested lines showed GUS expression in various rice organs (Ito et al. 2004).

Forward Screening

Using T-DNA-tagged lines, forward genetics have been performed. Because most knockout mutations are recessive, their phenotypes can be detected only in the progeny generations. Jeon and An (2001) have reported that commonly observed mutants were the dwarf (7.0%) and leaf pigment alterations (9.5%). Spotted leaves (1.0%) and leaf morphology mutants (1.2%) were also found. Likewise, Wu et al. (2003) have reported morphologically conspicuous mutations in about 7.5% of their 2679 T-DNA tagging lines. Screening 15,586 T-DNA tagged lines identified 81 (0.52%) lines that showed cold-responsive GUS activity (S.-C. Lee et al. 2004b). Among them, 16 lines displayed abscisic acid (ABA)-inducible GUS activity.

When co-segregation analyses of phenotypes with the insert DNA have been tested, many of the phenotypes were not caused by the inserts (<5%, unpublished data). This might be due to the tissue culture process of about 3 or 4 months, generating mutations not associated with T-DNA insertions (Phillips et al. 1994, Kaeppler et al. 2000). Retrotransposons, such as *Tos17* or *Karma*, and miniature inverted repeat transposable elements (MITEs), e.g. *mPing* or *Pong*, can act as mutagens (Hirochika et al. 1996, Jiang et al. 2003, Kikuchi et al. 2003, Komatsu et al. 2003, Nakazaki et al. 2003). In addition, small insertions, deletions and base substitutions may be induced in cultured cells (Miyao et al. 2003). Nonetheless, several genes such as *OsMADS50*, the magnesium chelatase gene and *OsDMKT1* have been isolated from T-DNA-tagged lines using forward genetic screening (Jung et al. 2003, S. Lee et al. 2004, S.-C. Lee et al. 2004b). To enhance the efficiency of forward genetics in rice, it will be necessary to develop a new transformation technique skipping the tissue culture procedure.

DNA Pool Screening for Reverse Genetics

Although a large number of rice genes have now been identified, via their sequence homology to pre-identified genes, nevertheless, their functional roles in rice are mostly unknown. DNA chip and proteome analyses based on genome information provide valuable information concerning expression, which accelerates the accumulation of basic functional knowledge at the genomic level. However, because mutants are essential for elucidating target gene functions, reverse genetics is an excellent method for studying function relatively quickly. In these approaches, mutations in a target gene are identified, then their effects are investigated in the mutant lines.

A polymerase chain reaction (PCR)-based screening method has been derived for identifying insertional mutants in a given gene (Krysan et al. 1999, Sato et al. 1999). Knockout mutants are screened with a set of primers: a gene-specific primer and a primer located near the end of the insert. The templates used with this method are pools of DNA prepared from 100–1,000 lines that contain T-DNA insertions. PCR products are run on an agarose gel, blotted onto a membrane, and hybridized with a gene-specific probe. Once a positive band is identified, DNAs from corresponding subpools and individuals are screened sequentially for identification of a line with the T-DNA insert. The DNA fragment flanking the insert element is then amplified and its sequence determined.

Because this approach is labor-intensive and time-consuming, DNA pools for a large number of lines are commonly used with this reverse genetics tool. In *Arabidopsis*, pools of 1,000–5,000 lines have been efficiently utilized for identifying insertional mutations (Winkler et al. 1998, Krysan et al. 1999, Meissner et al. 1999, Rios et al. 2002). Because the rice genome is about four times larger than that of *Arabidopsis*, a pool of 1,000 lines is considered sufficient for PCR-based screening of its knockout mutations (Lee et al. 2003). Rice also has a higher amount of GC than does *Arabidopsis*, especially at the 5′ ends of its genes (Carels and Bernardi 2000, Yu et al. 2002). These high GC regions are difficult to amplify under normal PCR conditions. Therefore, use of a GC buffer and adjustment of the annealing temperature can improve the screening efficiency of such GC-rich genes (Hengen 1997, Henke et al. 1997, Lee et al. 2003). Nevertheless, several primers must be examined when analyzing large genes because PCR efficiency is quite low when the product is >1 kb.

This screening strategy has been utilized successfully in *Arabidopsis*, petunia and maize (Koes et al. 1995, Mena et al. 1996, Krysan et al. 1999, Parinov and Sundaresan 2000). For example, 17 insertions in 63 genes involved in signal transduction and ion transport, 47 insertions in 36 members of the *R2R3 MYB* gene family, and 22 mutations in 70 *P450* genes have been isolated from *Arabidopsis* (Winkler et al. 1998, Krysan et al. 1999, Meissner et al. 1999). In petunia, a DNA pool has been generated from 1,000 individual plants with highly active *dTph1* elements; PCR screening has resulted in the identification of *dTph1* insertions for 10 genes (Koes et al. 1995). Reverse genetics screening has also led to the isolation of a transposon-induced mutation in *ZAG1*, the maize homolog of *AGAMOUS* (Mena et al. 1996). In rice, PCR screening for 12 MADS box genes of DNA pools prepared from 21,049 T-DNA-tagged lines has enabled the identification of five insertions in four genes (Lee et al. 2003). The DNA pool size in the laboratory of An has been increased to 60,000 lines and the success rate is approximately 50%.

Establishment of a Tag End Sequence (TES) Database

Despite the usability of PCR screening of DNA pools, it has several limitations: (i) a finite amount of DNA pools; (ii) a labor- and time-intensive screening process; and (iii) limited information about T-DNA integration sites. Therefore, establishment of a database for T-DNA insertion sites should facilitate the use of T-DNA-tagged lines. Tag end sequences (TES) that flank insert elements have been obtained by various methods, such as thermal asymmetric interlaced PCR (TAIL PCR),

Table 2 Insertion characteristics of T-DNA, *Tos17* and *Ds* in rice

	T-DNA	*Tos17*	*Ds*
Randomness	High	Low	?
Preferred chromosome	1, 2, 3, 6	1, 2, 3, 6	1, 4, 7
Avoided chromosome	9, 10, 12	8, 9, 10	9, 11, 12
GC content around insertion sites	Normal	Narrow GC content region	?
Preferred area (genic/intergenic)	Genic	Genic	Genic
Preferred region (exon/intron)	No preference	No preference	No preference
Preferred gene class	No preference	Kinase/disease	?

inverse PCR (iPCR) or adaptor ligation PCR (Ochman et al. 1988, Triglia et al. 1988, Rosenthal and Jones 1990, Liu and Whittier 1995, Siebert et al. 1995). Large-scale application of this strategy requires considerable effort but, once established, these databases can be easily shared with other scientists, facilitating the distribution of mutant materials and the analysis of gene functioning (Parinov et al. 1999, Tissier et al. 1999, Parinov and Sundaresan 2000, An et al. 2003). With sequencing of their entire genomes nearly complete, flanking sequence databases for rice and *Arabidopsis* will become powerful tools for systematically analyzing the functions of a large number of their genes (Walbot 1992, Walbot 2000, Parinov and Sundaresan 2000, Kumar and Hirochika 2001, Pan et al. 2003). Databases for T-DNA and *Ds* transposon insertion site sequences have already been developed for *Arabidopsis* (Parinov et al. 1999, Tissier et al. 1999, Ortega et al. 2002, Sessions et al. 2002). In maize, the DNAs adjacent to transposed *Ac* elements have also been isolated and sequenced (Cowperthwaite et al. 2002).

In rice, TES databases for T-DNA insertional lines have also been created. An et al. (2003) and Ryu et al. (2004) have reported the establishment of a database of 4,783 TES (www.postech.ac.kr/life/pfg/risd), while Chen et al. (2003) have described one comprising 1,009 TES (www.genomics.zju.edu.cn/ricetdna). The TES database is rapidly growing, and its public site currently contains 24,299 TES obtained from the An laboratory (www.postech.ac.kr/life/pfg/risd). Recently, Sallaud and colleagues have analyzed 7,480 TES from T-DNA-tagged lines (Sallaud et al. 2004), and *Tos17* and *Ds* tag databases have also been constructed in rice (Miyao et al. 2003, Kolesnik et al. 2004).

TES analyses not only provide researchers with information about disrupted genes, but also reveal the distribution of insert elements in plant chromosomes. In rice, three kinds of insertion elements (T-DNA, *Tos17* and *Ds*) have been employed for generating large-scale mutant pools (An et al. 2003, Chen et al. 2003, Miyao et al. 2003, Kolesnik et al. 2004, Sallaud et al. 2004). Each element manifests characteristic insertion patterns (Table 2). T-DNA insertions appear to be less prone to hot spots when compared with those of *Tos17* and *Ds*, which are aggregated in hot spots (Miyao et al. 2003, Kolesnik et al. 2004, Sallaud et al. 2004). T-DNA insertion density is higher in four chromosomes (i.e. 1, 2, 3, and 6) but lower in three chromosomes (9, 10 and 12) when compared with the others (Sallaud et al. 2004). Similar patterns have been observed in *Tos17* insertion analyses, perhaps because of the difference in gene density for each chromosome (Sallaud et al. 2004). Overall GC content at the T-DNA insertion sites is close to that measured for the entire rice genome (An et al. 2003, Chen et al. 2003), whereas *Tos17* target regions have a narrow pattern of GC content distribution (Miyao et al. 2003).

Furthermore, *Tos17* and *Ds* are inserted into genic regions (from ATG to the stop codon) more frequently than into intergenic regions (Miyao et al. 2003, Kolesnik et al. 2004). Apparently, T-DNA has a higher insertion ratio near the start ATG codon (An et al. 2003, Chen et al. 2003, Sallaud et al. 2004). In the case of *Arabidopsis*, one can observe a tendency for T-DNA insertions in the intergenic regions (Sessions et al. 2002, Alonso et al. 2003). Regarding exon/intron preference, T-DNA does not exhibit the difference that *Ds* does (An et al. 2003, Kolesnik et al. 2004, Sallaud et al. 2004). Chen et al. (2003) have reported that T-DNA insertions are relatively dense in introns (more than twice that for exons); however, this conclusion needs to be verified with additional TES data. In contrast, other researchers have found that T-DNA insertions are almost equally distributed between exons and introns (An et al. 2003, Sallaud et al. 2004). When functional classification of genes tagged by T-DNA has been performed, T-DNA insertion has not been found to be biased toward a particular type of gene, whereas *Tos17* insertions do occur preferentially in the ‘kinase’ and ‘disease resistance’ classes (An et al. 2003, Miyao et al. 2003).

Although these three insertion elements differ in several ways, they have common advantages: (i) genome-wide distribution; (ii) preferred insertion into gene-rich regions; and (iii) low frequencies of integration within repetitive, or pericentromeric, regions. Along with possessing the most random distribution patterns, these advantages make T-DNA an ideal insertional mutagen for rice functional genomics.

Table 3 shows that the chance of finding T-DNA knockout genes from the TES database established in the laboratory of An is currently 35%. Approximately 43% of rice genes have putative activation taggings in the database. The laboratory

Table 3 Knockout and putative activation tagging mutants present in the TES database developed in the An laboratory

Gene	Gene number	Knockout gene [a] (mutant number)	Activation gene [b] (mutant number)
BAK1-like kinase	13	2 (2)	7 (13)
BRI1-like kinase	28	9 (10)	10 (20)
CLAVATA1-like kinase	27	9 (14)	10 (14)
ERECTA-like kinase	5	1 (1)	4 (6)
MAPK	15	3 (4)	5 (7)
MAPKK	9	3 (3)	3 (4)
MAPKKK	15	2 (4)	6 (10)
Genes involved in starch biosynthesis from sucrose	31	15 (26)	12 (25)
Omega-3 fatty acid desaturase	5	2 (3)	1 (3)
Lipoxygenase	4	1 (1)	0 (0)
CER1-like	8	6 (8)	8 (12)
Double AP2 domain	23	14 (29)	18 (39)
bHLH	166	58 (71)	84 (123)
MADS	78	25 (40)	16 (20)
CONSTANS-like	16	5 (8)	2 (2)
FT-like	12	2 (3)	0 (0)
Phytochrome	3	3 (4)	1 (1)
MLO-like	14	6 (7)	9 (12)
HMA	9	5 (8)	6 (11)
ZIP	12	4 (5)	5 (7)
YS1-like	18	6 (6)	12 (25)
Total	511	181	219

[a] T-DNA insertion within 300 bp outside from the start ATG and stop codon of the putative genes.
[b] Within 10 kb from the start ATG codon.

plans to establish and publicize at least 50,000 independent TES within the next 2 years.

Reverse Genetics Study Using T-DNA-tagged Lines

The reverse genetics approach has been adapted successfully since it was first introduced by McKinney et al. (1995) in their study to identify insertions in two members of the actin gene family from a T-DNA population of 5,300 *Arabidopsis* transformants (about 8000 inserts). The genes targeted in this approach belong to almost all functional categories, such as those for metabolism (Perrin et al. 2003, Fulda et al. 2004), cell division (Couteau et al. 1999, Siddiqui et al. 2003), transcription (Gualberti et al. 2002, S.-C. Lee et al. 2004b), transporters (Kaiser et al. 2002, E.K. Lee et al. 2004), cell structure (Li et al. 2004, Usadel et al. 2004) and signal transduction (Pischke et al. 2002, Leonhardt et al. 2004). Most of these studies have been conducted in *Arabidopsis* because of the large, publicly available T-DNA insertion libraries and TES databases. Now that those libraries and databases are being rapidly established for rice, reverse genetics will be widely used in this important crop species as well.

Obstacles in Reverse Genetics

The phenotypic alterations observed in a T-DNA-tagged line are not necessarily due to the insertional mutation caused by the element. This is because T-DNA often inserts into more than one locus in a particular chromosome. Alternatively, other elements, such as endogenous transposons, might be involved in the mutant phenotype. Tissue culturing often causes point mutations as well as small deletions and insertions. Therefore, it is necessary to confirm those observations by analyzing additional lines that contain a mutation in the target gene. Large databases for knockout mutants, not only from T-DNA but also from transposons, are extremely important for identifying multiple alleles in a given gene. DNA pools are also valuable in enriching alleles. If an allele is not present, antisense or RNAi approaches to reducing gene expression are useful. However, these alternative methods may affect the expression of genes that are structurally similar to the target gene. In addition, a small amount of expression persists in antisense or RNAi plants, a level that is often sufficient to produce a normal phenotype. In these cases, complementation of the mutant with wild-type cDNA or a genomic clone can prove whether the mutant phenotype is indeed due to the T-DNA insertion.

One major obstacle in the reverse genetics approach is gene duplication, in which redundancy may result in no phenotypic alteration. Researchers with the rice genome project have predicted that a large number of its genes are duplicated or polyploidized (Goff et al. 2002, Yu et al. 2002). Therefore, double or multiple mutations in a group of related genes are necessary for observing mutant phenotypes (Liljegren et al. 2000, Pelaz et al. 2000). To this end, numerous tagging populations and TES databases have been generated, and Web sites for those databases are being constructed (Hirochika et al. 2004). These databases will facilitate the identification of individual mutant lines of related genes.

Maintenance and distribution of mutant seeds are also major problems. Seed from primary transgenic plants should be amplified before distribution because those types of plants produce only a few seeds (Jeon et al. 2000). Because T-DNA-tagged lines contain transgenic elements, e.g. reporter genes and selectable markers, they are considered genetically modified (GM) plants. Therefore, only limited space is available for seed amplification. The resultant seed must then be stored in a controlled environment of low temperature and humidity because viability is rapidly reduced under warm, humid conditions. To ensure the safety of the seed during emergency situations, collections should be kept in multiple locations. Ideally, in addition to holding the original mutant stock in developed laboratories, a second copy should be maintained by a well-established international organization, such as the International Rice Research Institute, while a third copy is being preserved at the governmental genetic resource center where the lines were first generated. As a final obstacle, the sharing of research materials is often slowed because seed distribution to other countries requires a long quarantine process.

Future Projections

T-DNA insertional mutant libraries are now publicly available (Table 1). Nevertheless, for saturation mutagenesis, a larger number of mutants are needed. Transposon tagging will provide a good complementation to T-DNA insertional mutagenesis, especially for recalcitrant cultivars. The generation of TES databases for mutant libraries should accelerate the use of such lines. It will also be valuable to develop a database for phenotypic alterations in mutant lines. As has been done with *Arabidopsis* (Ichikawa et al. 2003), the evaluation of tagged lines should include the assessment of traits for any visible phenotypes, such as abnormalities in morphology, growth rate, plant color, flowering time and fertility.

Acknowledgments

We thank Priscilla Licht for critical editing of the manuscript. The gene tagging project for functional genomics of rice in the An laboratory is supported, in part, by grants from the Crop Functional Genomic Center, the 21st Century Frontier Program (CG-1111), from the Biogreen 21 Program, Rural Development Administration and from POSCO.

References

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., et al. (2003) *Science* 301: 653–657.

Alvarado, M.C., Zsigmond, L.M., Kovacs, I., Cseplo, A., Koncz, C. and Szabados, L.M. (2004) *Plant Physiol.* 134: 18–27.

An, S., Park, S., Jeong, D.-H., Lee, D.Y., Kang, H.G., et al. (2003) *Plant Physiol.* 133: 2040–2047.

Azpiroz-Leehan, R. and Feldmann, K.A. (1997) *Trends Genet.* 13: 152–156.

Arabidopsis Genome Initiative (2000) *Nature* 408: 796–815.

Barakat, A., Gallois, P., Raynal, M., Mestre-Ortega, D., Sallaud, C., Guiderdoni, E., Delseny, M. and Bernardi, G. (2000) *FEBS Lett.* 471: 161–164.

Bouche, N. and Bouchez, D. (2001) *Curr. Opin. Plant Biol.* 4: 111–117.

Brand, A.H. and Perrimon, N. (1993) *Development* 118: 401–415.

Campisi, L., Yang, Y.Z., Yi, Y., Heilig, E., Herman, B., Cassista, A.J., Allen, D.W., Xiang, H.J. and Jack, T. (1999) *Plant J.* 17: 699–707.

Carels, N. and Bernardi, G. (2000) *Genetics* 154: 1819–1825.

Chen, S., Jin, W., Wang, M., Zhang, F., Zhou, J., Jia, Q., Wu, Y., Liu, F. and Wu, P. (2003) *Plant J.* 36: 105–113.

Chin, H.G., Choe, M.S., Lee, S.H., Park, S.H., Park, S.H., et al. (1999) *Plant J.* 19: 615–623.

Chuang, C.F. and Meyerowitz, E.M. (2000) *Proc. Natl Acad. Sci. USA* 97: 4985–4990.

Cornejo, M.J., Luth, D., Blankenship, K.M., Anderson, O.D. and Blechl, A.E. (1993) *Plant Mol. Biol.* 23: 567–581.

Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D. and Doutriaux, M.P. (1999) *Plant Cell* 11: 1623–1634.

Cowperthwaite, M., Park, W., Xu, Z., Yan, X., Maurais, S.C. and Dooner, H.K. (2002) *Plant Cell* 14: 713–726.

Feldmann, K.A. (1991) *Plant J.* 1: 71–82.

Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G. et al. (2002) *Nature* 420: 316–320.

Fulda, M., Schnurr, J., Abbadi, A., Heinz, E. and Browse, J. (2004) *Plant Cell* 16: 394–405.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R. et al. (2002) *Science* 296: 92–100.

Greco, R., Ouwerkerk, P.B., de Kam, R.J., Sallaud, C., Favalli, C., Colombo, L., Guiderdoni, E., Meijer, A.H., Hoge Dagger, J.H. and Pereira, A. (2003) *Theor. Appl. Genet.* 108: 10–24.

Gualberti, G., Papi, M., Bellucci, L., Ricci, I., Bouchez, D., Camilleri, C., Costantino, P. and Vittorioso, P. (2002) *Plant Cell* 14: 1253–1263.

Hanin, M. and Paszkowski, J. (2003) *Plant Biol.* 6: 157–162.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., et al. (1998) *Genetics* 148: 479–494.

Haseloff, J., Siemering, K.R., Prasher, D.C. and Hodge, S. (1997) *Proc. Natl Acad. Sci. USA* 94: 2122–2127.

Hengen, P.N. (1997) *Trends Biochem. Sci.* 22: 225–226.

Henikoff, S., Till, B.J. and Comai, L. (2004) *Plant Physiol.* 135: 630–636.

Henke, W., Herdel, K., Jung, K., Schnorr, D. and Loening, S.A. (1997) *Nucleic Acids Res.* 25: 3957–3958.

Hiei, Y., Ohta, S., Komari, T. and Kumashiro, T. (1994) *Plant J.* 6: 271–282.

Hirochika, H., Guiderdoni, E., An, G., Hsing, Y.I., Eun, M.Y., et al. (2004) *Plant Mol. Biol.* 54: 325–334.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H. and Kanda, M. (1996) *Proc. Natl Acad. Sci. USA* 93: 7783–7788.

Ichikawa, T., Nakazawa, M., Kawashima, M, Muto, S., Gohda, K., et al. (2003) *Plant J.* 36: 421–429.

Ito, Y., Eiguchi, M. and Kurata, N. (2004) *Mol. Genet. Genomics* 271: 639–650.

Ito, T. and Meyerowitz, E.M. (2000) *Plant Cell* 12: 1541–1550.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) *EMBO J.* 6: 3901–3907.

Jeon, J.S. and An, G. (2001) *Plant Sci.* 161: 211–219.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.-H., et al. (2000) *Plant J.* 22: 561–570.

Jeong, D.-H., An, S., Kang, H.-G., Moon, S., Han, J.-J., Park, S., Lee, H.S., An, K. and An, G. (2002) *Plant Physiol.* 130: 1636–1644.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R. and Wessler, S.R. (2003) *Nature* 421: 163–167.

Jung, K.H., Hur, J., Ryu, C.H., Choi, Y., Chung, Y.Y., Miyao, A., Hirochika, H. and An, G. (2003) *Plant Cell Physiol.* 33: 463–472.

Kaeppler, S.M., Kaeppler, H.F. and Rhee, Y. (2000) *Plant Mol. Biol.* 43: 179–188.

Kaiser, B.N., Rawat, S.R., Siddiqui, M.Y., Masle, J. and Glass, A.D. (2002) *Plant Physiol.* 130: 1263–1275.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J. and Weigel, D. (1999) *Science* 286: 1962–1965.

Kikuchi, K., Terauchi, K., Wada, M. and Hirano, H.Y. (2003) *Nature* 421: 167–170.

Kirik, V., Simon, M., Huelskamp, M. and Schiefelbein, J. (2004) *Dev. Biol.* 268: 506–513.

Klimyuk, V.I., Nussaume, L., Harrison, K. and Jones, J.D.G. (1995) *Mol. Gen. Genet.* 249: 357–365.

Koes, R., Souer, E., van Houwelingen, A., Mur, L., Spelt, C., et al. (1995) *Proc. Natl Acad. Sci. USA* 92: 8149–8153.

Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C.S., Jiang, S., Ramamoorthy, R., Cai, M., Ma, Z.G., Sundaresan, V. and Ramachandran, S. (2004) *Plant J.* 37: 301–314.

Komatsu, M., Shimamoto, K. and Kyozuka, J. (2003) *Plant Cell* 15: 1934–1944.

Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) *Plant Cell* 11: 2283–2290.

Kumar, A. and Hirochika, H. (2001) *Trends Plant Sci.* 6: 127–134.

LeClere, S. and Bartel, B. (2001) *Plant Mol. Biol.* 46: 695–703.

Lee, E.K., Kwon, M., Ko, J.H., Yi, H., Hwang, M.G., Chang, S. and Cho, M.H. (2004) *Plant Physiol.* 134: 528–538.

Lee, H., Suh, S.S., Park, E., Cho, E., Ahn, J.H., Kim, S.G., Lee, J.S., Kwon, Y.M. and Lee, I. (2000) *Genes Dev.* 14: 2366–2376.

Lee, S., Jeon, J.S., Jung, K.H. and An, G. (1999) *J. Plant Biol.* 42: 310–316.

Lee, S., Kim, J., Han, J.J., Han, M.J. and An, G. (2004) *Plant J.* 38: 754–764.

Lee, S.-C., Kim, J.-Y., Kim, S.-H., Kim, S.-J., Lee, K., Han, S.-K., Choi, H.-S., Jeong D.-H., An, G. and Kim, S.-R. (2004b) *Plant Sci.* 166: 69–79.

Lee, S., Kim, J., Son, J.S., Nam, J., Jeong, D.-H., et al. (2003) *Plant Cell Physiol.* 244: 1403–1411.

Leonhardt, N., Kwak, J.M., Robert, N., Waner, D., Leonhardt, G. and Schroeder, J.I. (2004) *Plant Cell* 16: 596–615.

Li, X., Cordero, I., Caplan, J., Molhoj, M. and Reiter, W.D. (2004) *Plant Physiol.* 134: 940–950.

Liljegren, S.J., Ditta, G.S., Eshed, Y., Savidge, B., Bowman, J.L. and Yanofsky, M.F. (2000) *Nature* 404: 766–770.

Liu, Y.G. and Whittier, R. (1995) *Genomics* 25: 674–681.

Matsuhara, S., Jingu, F., Takahashi, T. and Komeda, Y. (2000) *Plant J.* 22: 79–86.

McElroy, D., Zhang, W., Cao, J. and Wu, R. (1990) *Plant Cell* 2: 163–171.

McKinney, E.C., Ali, N., Traut, A., Feldmann, K.A., Belostotsky, D.A., McDowell, J.M. and Meagher, R.B. (1995) *Plant J.* 8: 613–622.

Meissner, R., Chague, V., Zhu, Q., Emmanuel, E., Elkind, Y. and Levy, A.A. (2000) *Plant J.* 22: 265–274.

Meissner, R.C., Jin, H., Cominelli, E., Denekamp, M., Fuertes, A., et al. (1999) *Plant Cell* 11: 1827–1840.

Mena, M., Ambrose, B.A., Meeley, R.B., Briggs, S.P., Yanofsky, M.F. and Schmidt, R.J. (1996) *Science* 274: 1537–1540.

Millar, A.J., Short, S.R., Chua, N.H. and Kay, S.A. (1992) *Plant Cell* 4: 1075–1087.

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K. and Hirochika, H. (2003) *Plant Cell* 15: 1771–1780.

Nagamura, Y., Inoue, T., Antonio, B.A., Shimano, T., Kajiya, H., et al. (1995) *Breed. Sci.* 45: 373–376.

Nakazaki, T., Okumoto, Y., Horibata, A., Yamahira, S., Teraishi, M., Nishida, H., Inoue, H. and Tanisaka, T. (2003) *Nature* 421: 170–172.

Neff, M.M., Nguyen, S.M., Malancharuvil, E.J., Fujioka, S., Noguchi, T., Seto, H., Tsubuki, M., Honda, T., Takatsuto, S., Yoshida, S. and Chory, J. (1999) *Proc. Natl Acad. Sci. USA* 96: 15316–15323.

Ochman, H., Gerber, A.S. and Hartl, D.L. (1988) *Genetics* 120: 621–623.

Odell, J.T., Nagy, F. and Chua, N. (1985) *Nature* 313: 810–812.

Ortega, D., Raynal, M., Laudie, M., Llauro, C., Cooke, R., et al. (2002) *C. R. Biol.* 325: 773–780.

Pan, X., Liu, H., Clarke, J., Jones, J., Bevan, M. and Stein, L. (2003) *Nucleic Acids Res.* 31: 1245–1251.

Pang, S.Z., DeBoer, D.L., Wan, Y., Ye, G., Layton, J.G., Neher, M.K., Armstrong, C.L., Fry, J.E., Hinchee, M.A. and Fromm, M.E. (1996) *Plant Physiol.* 112: 893–900.

Parinov, S., Sevugan, M., De, Y., Yang, W.C., Kumaran, M. and Sundaresan, V. (1999) *Plant Cell* 11: 2263–2270.

Parinov, S. and Sundaresan, V. (2000) *Curr. Opin. Biotechnol.* 11: 157–161.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F. (2000) *Nature* 405: 200–203.

Perrin, R.M., Jia, Z., Wagner, T.A., O'Neill, M.A., Sarria, R., York, W.S., Raikhel, N.V. and Keegstra, K. (2003) *Plant Physiol.* 132: 768–778.

Phelps, B.B. and Brand, A.H. (1998) *Methods Enzymol.* 14: 367–379.

Phillips, R.L., Kaeppler, S.M. and Olhoft, P. (1994) *Proc. Natl Acad. Sci. USA* 91: 5222–5226.

Pischke, M.S., Jones, L.G., Otsuga, D., Fenandez. D.E., Drews, G.N. and Sussman, M.R. (2002) *Proc. Natl.Acad. Sci. USA* 99: 15800–15805.

Rios, G., Lossow, A., Hertel, B., Breuer, F., Schaefer, S., et al. (2002) *Plant J.* 32: 243–253.

Robertson, D. (2004) *Annu. Rev. Plant Biol.* 55: 495–519.

Rodermel, S.R., Abbott, M.S. and Bogorad, L. (1988) *Cell* 55: 673–681.

Rosenthal, A. and Jones, D.S. (1990) *Nucleic Acids Res.* 18: 3095–3096.

Ryu, C.-H., You, J.-H., Kang, H.-G., Hur, J., Kim, Y.-H., Han, M.-J., An, K., Chung, B.-C., Lee, C.-H. and An, G. (2004) *Plant Mol. Biol.* 54: 489–502.

Sallaud, C., Gay, C., Larmande, P., Bes, M., Piffanelli, P., et al. (2004) *Plant J.* 39: 450–464.

Sallaud, C., Meynard, D., van Boxtel, J., Gay, C., Bes, M., et al. (2003) *Theor. Appl. Genet.* 106: 1396–1408.

Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T. et al. (2002) *Nature* 420: 312–316.

Sato, Y., Sentoku, N., Miura, Y., Hirochika, H., Kitano. H. and Matsuoka, M. (1999) *EMBO J.* 18: 992–1002.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., et al. (2002) *Plant Cell* 14: 2985–2994.

Sha, Y., Li, S., Pei, Z., Luo, L., Tian, Y. and He, C. (2004) *Theor. Appl. Genet.* 108: 306–314.

Siddiqui, N.U., Stronghill, P.E., Dengler, R.E., Hasenkampf, C.A. and Riggs, C.D. (2003) *Development* 130: 3283–3295.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A. and Lukyanov, S.A. (1995) *Nucleic Acids Res.* 23: 1087–1088.

Springer, P.S. (2000) *Plant Cell* 12: 1007–1020.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D., Dean, C., Ma, H. and Martienssen, R. (1995) *Genes Dev.* 9: 1797–1810.

Sundaresan, V. (1996) *Trends Plant Sci.* 1: 184–191.

Szabados, L., Kovacs, I., Oberschall, A., Abraham, E., Kerekes, I., et al. (2002) *Plant J.* 32: 233–242.

Tani, H., Chen, X., Nurmberg, P., Grant, J.J., SantaMaria, M., Chini, A., Gilroy, E., Birch, P.R. and Loake, G.J. (2004) *Funct. Integr. Genomics* 4: 258–266.

Thompson, J.F., Hayes, L.S. and Lloyd, D.B. (1991) *Gene* 103: 171–177.

Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G. and Jones, J.D. (1999) *Plant Cell* 11: 1841–1852.

Topping, J.F. and Lindsey, K. (1997) *Plant Cell* 9: 1713–1725.

Triezenberg, S.J., Kingsbury, R.C. and McKnight, S.L. (1988) *Gene Dev.* 2: 718–729.

Triglia, T., Peterson, M.J. and Kemp, D.J. (1988) *Nucleic Acids Res.* 16: 8186.

Usadel, B., Kuschinsky, A.M., Rosso, M.G., Eckermann, N. and Pauly, M. (2004) *Plant Physiol.* 134: 286–295.

van der Graaff, E., Dulk-Ras, A.D., Hooykaas, P.J. and Keller, B. (2000) *Development* 127: 4971–4980.

Walbot, V. (1992) *Mol. Gen. Genet.* 234: 353–360.

Walbot, V. (2000) *Curr. Opin. Plant Biol.* 3: 103–107.

Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., et al. (2000) *Plant Physiol.* 122: 1003–1013.

Wilson, W.A., Harrington, S.E., Woodman, W.L., Lee, M., Sorrells, M.E. and McCouch, S.R. (1999) *Genetics* 153: 453–473.

Winkler, R.G., Frank, M.R., Galbraith, D.W., Feyereisen, R. and Feldmann, K.A. (1998) *Plant Physiol.* 118: 743–750.

Wu, C., Li, X., Yuan, W., Chen, G., Kilian, A., Li, J., Xu, C., Li, X., Zhou, D.X., Wang, S. and Zhang, Q. (2003) *Plant J.* 35: 418–427.

Yin, Z. and Wang, G.-L. (2000) *Theor. Appl. Genet.* 100: 461–470.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S. et al. (2002) *Science* 296: 79–92.

Zuo, J., Niu, Q.W., Frugis, G. and Chua, N.H. (2002) *Plant J.* 30: 349–359.

(Received October 28, 2004; Accepted November 7, 2004)