

TECHNICAL ADVANCE

Rapid identification of *Arabidopsis* insertion mutants by non-radioactive detection of T-DNA tagged genes

Gabino Ríos¹, Andrea Lossow¹, Britta Hertel¹, Frank Breuer¹, Sabine Schaefer¹, Melanie Broich¹, Tatjana Kleinow¹, Ján Jásik¹, Jochen Winter¹, Alejandro Ferrando¹, Rosa Farrás¹, Mireia Panicot¹, Rossana Henriques¹, Jean-Baptist Mariaux¹, Attila Oberschall¹, Gergely Molnár¹, Kenneth Berendzen¹, Vijaya Shukla¹, Marcel Lafos¹, Zsuzsanna Koncz¹, George P. Rédei², Jeff Schell¹ and Csaba Koncz^{1,*}

¹Max-Planck Institut für Züchtungsforschung, Carl-von-Linné-Weg 10, D-59829 Köln, Germany, and

²3005 Woodbine Ct, Columbia, MO 65203-0906, USA.

Received 24 April 2002; revised 17 June; accepted 19 June 2002.

*For correspondence (fax + 49 221 5062213; e-mail koncz@mpiz-koeln.mpg.de).

Summary

To assist in the analysis of plant gene functions we have generated a new *Arabidopsis* insertion mutant collection of 90 000 lines that carry the T-DNA of *Agrobacterium* gene fusion vector pPCV6NFHyg. Segregation analysis indicates that the average frequency of insertion sites is 1.29 per line, predicting about 116 100 independent tagged loci in the collection. The average T-DNA copy number estimated by Southern DNA hybridization is 2.4, as over 50% of the insertion loci contain tandem T-DNA copies. The collection is pooled in two arrays providing 40 PCR templates, each containing DNA from either 4000 or 5000 individual plants. A rapid and sensitive PCR technique using high-quality template DNA accelerates the identification of T-DNA tagged genes without DNA hybridization. The PCR screening is performed by agarose gel electrophoresis followed by isolation and direct sequencing of DNA fragments of amplified T-DNA insert junctions. To estimate the mutation recovery rate, 39 700 lines have been screened for T-DNA tags in 154 genes yielding 87 confirmed mutations in 73 target genes. Screening the whole collection with both T-DNA border primers requires 170 PCR reactions that are expected to detect a mutation in a gene with at least twofold redundancy and an estimated probability of 77%. Using this technique, an M_2 family segregating a characterized gene mutation can be identified within 4 weeks.

Keywords: T-DNA, insertion mutagenesis, *Arabidopsis*, PCR screening, functional genomics.

Introduction

Identification of insertion and point mutations in all *Arabidopsis* genes is a major goal of the Multinational Coordinated *Arabidopsis* 2010 project on functional genomics. Currently, the genome project uses three different mutational approaches that are based on the identification of chemically induced mutations with the TILLING technology (Colbert *et al.*, 2001; McCallum *et al.*, 2000a; McCallum *et al.*, 2000b) and the isolation of insertion mutations by either transposons or the T-DNA of *Agrobacterium*. Transposon-tagged mutant populations carry insertions of the autonomous maize transposon *En* (ZIGIA *En/Spm* lines: Baumann *et al.*, 1998; Wisman *et al.*, 1998a; Wisman *et al.*, 1998b), and stabilized inserts of the *Suppressor-mutator* (SLAT-lines, Tissier *et al.*, 1999); *En-I* (ITS-lines, Aarts *et al.*, 1995; Speelman *et al.*, 1999; Speelman *et al.*, 2000); and *Ac/Ds*

(Bancroft and Dean, 1993; Bancroft *et al.*, 1992; Fedoroff and Smith, 1993; Long *et al.*, 1993; Long *et al.*, 1997) two-element transposon tagging systems. Mutations are also induced by the tobacco retrotransposon *Tto1* (Okamoto and Hirochika, 2000) and several other transposon constructs exploiting the enhancer-trap and Cre/Lox recombination technologies (Martienssen, 1998; Osborne *et al.*, 1995; Smith *et al.*, 1996; Sundaresan *et al.*, 1995). In addition to the first widely distributed T-DNA insertion mutant collections (Azpiroz-Leehan and Feldmann, 1997; Feldmann, 1991; Koncz *et al.*, 1992), new mutant populations are also available (with gene and promoter trap inserts) that drive specific expression of β -glucuronidase (GUS) and green fluorescent reporter proteins (GFP) (Devic *et al.*, 1995; Goddijn *et al.*, 1993; Kiegle *et al.*, 2000; Topping *et al.*,

1991); and with activator T-DNA tags that facilitate screening for dominant mutations (Weigel *et al.*, 2000). Saturation T-DNA mutagenesis is now performed using *in planta* transformation (Bechtold *et al.*, 1993; Clough and Bent, 1998) and exploited to identify gene mutations by direct sequencing of transposon and T-DNA insert junctions (Balzergue *et al.*, 2001; Liu *et al.*, 1995; Mathur *et al.*, 1998; Okamoto and Hirochika, 2000; Parinov *et al.*, 1999; Samson *et al.*, 2002; Speulman *et al.*, 1999; Tissier *et al.*, 1999; Yephremov and Saedler, 2000; <http://flagdb-genoplante-info.infobiogen.fr/projects/fst>; <http://genetrapp.cshl.org>; <http://signal.salk.edu/cgi-bin/tdnaexpress>; <http://www.jic.bbsrc.ac.uk/sainsbury-lab/jonathan-jones/SINS-database/sins.htm>; http://www.nadii.com/pages/collaborations/garlic_files/GarlicAnalysis.html). The availability of the complete *Arabidopsis* genome sequence (Arabidopsis Genome Initiative, 2000) also supports the wide ranging application of PCR-based reverse genetic approaches. Several T-DNA and transposon insertion mutant collections have been used to prepare arrays of pooled PCR DNA templates to screen for mutations in known genes and gene families (Bouché and Bouchez, 2001; Galbiati *et al.*, 2000; Krysan *et al.*, 1996; Krysan *et al.*, 1999; Mahalingam and Fedoroff, 2001; McKinney *et al.*, 1995; Meissner *et al.*, 1999; Parinov and Sundaresan, 2001; Thorneycroft *et al.*, 2001; Winkler *et al.*, 1998; Young *et al.*, 2001). However, the need to identify the insertions by Southern hybridization still limits the efficiency of PCR-based mutant screening techniques.

To contribute to the identification of gene knockouts, we describe here an improved PCR technique that allows simple detection of amplified T-DNA insert junctions by agarose gel electrophoresis, and subsequent isolation and sequencing of the PCR products. To exploit this method we have generated a new collection of 90 000 T-DNA tagged *Arabidopsis* lines by *in planta* transformation with the *Agrobacterium* gene fusion vector pPCV6NFHyg (Koncz *et al.*, 1989). Estimation of the average number of insertion loci and T-DNA copies indicates that the collection contains at least 116 100 independent insertion loci, over 50% of which carry concatenated T-DNA copies. To estimate the mutation frequency, 39 700 lines have been screened for T-DNA tags in 154 genes, which resulted in 87 confirmed knockouts in 73 genes. Based on this result, the technique is adjusted such that the entire population of 90 000 lines can be screened with 170 PCR reactions, leading to the identification of a sequenced T-DNA tag in a segregating M_2 mutant family within 4 weeks.

Results and discussion

Generation of *Arabidopsis* insertion mutant collection

To perform saturation T-DNA mutagenesis, *Arabidopsis* (Col-0) plants were transformed with *Agrobacterium* carry-

ing the gene fusion vector pPCV6NFHyg (Koncz *et al.*, 1989), using a modified vacuum infiltration protocol (Bechtold *et al.*, 1993; see Experimental procedures). The M_1 seed progeny were pooled, and aliquots germinated to determine the frequency of hygromycin-resistant, transformed seedlings. The results indicated that the M_1 pool contained about half a million transformed seeds. Subsequently, 90 000 hygromycin-resistant M_1 plants were planted in soil to collect an identical amount of leaf tissue in pools from each 100 plants. DNA was purified from the 900 pools and arranged in nine two-dimensional arrays (Figure 1).

The average number of independent insertion loci was estimated by determination of the segregation ratio of hygromycin-resistant to sensitive plants in 100 randomly selected M_2 families (Table 1). The majority of M_2 families (64%) segregated the T-DNA-encoded hygromycin-resistance marker at a ratio of 3:1, whereas 15:1 and 63:1 segregation ratios were observed in 16 and 2% of examined families, respectively. Lines showing 2:1 segregation and a significant deviation from the 15:1 ratio (indicating two linked inserts) were also observed. About 6% of M_2 families segregated in an exceptional manner, suggesting the potential occurrence of chromosomal rearrangements that may accompany the T-DNA integration process (Tax and Vernon, 2001). Excluding these lines, the average number of insertion loci was estimated as 1.29 per plant, similar to the values observed for other *Arabidopsis* T-DNA insertion mutant collections (1.47 and 1.53; Forsthoefel *et al.*, 1992). This suggested that the collection of 90 000 lines carried about 116 100 T-DNA insertion loci. To estimate the average number of T-DNA copies in the insertion loci, Southern DNA hybridization analysis was performed with randomly chosen plants. Figure 2(a) shows a representative hybridization of DNAs from 24 plants with the *Bam*HI–*Xba*I right-border fragment of pPCV6NFHyg that carries a promoterless *aph(3'III)* gene used for identification of plant gene fusions with the kanamycin-resistance marker (Koncz *et al.*, 1989). The hybridization results indicated that the average T-DNA insert number was about 2.4 per plant and that >50% of the insertion loci carried tandem T-DNA repeats, which occur with a similar frequency also in other T-DNA insertion mutant populations (Castle *et al.*, 1993; Feldmann and Marks, 1987).

Optimization of long-range PCR for sensitive detection of T-DNA insertions

To identify T-DNA tags in known genes, the PCR-based reverse-genetic screens utilize T-DNA end primers that are directed towards the left and right border repeats, in combination with 5' and 3' gene-specific primers oriented towards the coding regions of genes (Bouchez and Höfte, 1998; Figure 1). Two out of four possible combinations of gene-specific and T-DNA end primers are thus expected to

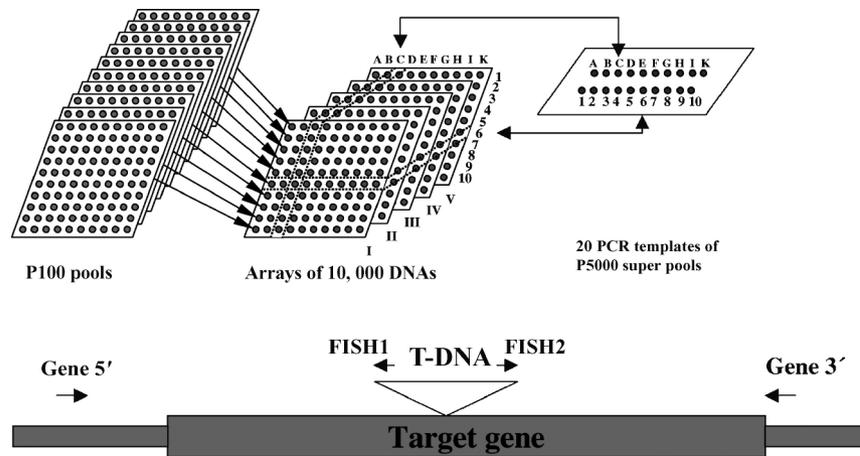


Figure 1. Construction of DNA arrays for PCR screening the insertion mutant collection.

Upper section: DNA samples of P100 pools, each representing 100 individual plants, were arranged in 10×10 arrays (labelled with roman numbers). Samples from the corresponding rows and columns of five arrays were pooled to generate number- and letter-coded P5000 super-pools. P4000 super-pools were obtained similarly.

Lower section: to screen for insertions in genes, the left (FISH1) or right (FISH2) T-DNA end primers were combined with either 5' or 3' gene-specific primers to screen each P4000 and P5000 super-pools with four PCR reactions. The first round of screening required 160 PCR reactions.

yield PCR products if an insertion was located in the target gene. To screen for T-DNA-induced mutations in our collection, two primers were designed for the left (FISH1) and right (FISH2) T-DNA ends, based on known sequence of the pPCV6NFHyg T-DNA (http://seeds.nottingham.ac.uk/Nasc/information/6nfhyg_map.lasso). The PCR reactions were optimized with three different mutant lines that carried characterized T-DNA tags in the *PRL1* (Németh *et al.*, 1998); *PAD1* (Farrás *et al.*, 2001; Tatjana Kleinow, unpublished results); and *At1g75960* genes (Gergely Molnár, unpublished results), using different thermotolerant DNA polymerases. The highest sensitivity of detection was obtained with the Takara LA Taq enzyme when the reactions contained $0.3\text{--}5\text{ ng } \mu\text{l}^{-1}$ template DNA in which the *prl1* DNA was diluted 1000-fold with DNA from wild-type Col-0 plants (data not shown). Testing different annealing temperatures indicated that the background was reduced to

minimal when annealing was performed at 68°C , identical to the recommended extension temperature for the LA Taq enzyme. Titration of the optimal number of PCR cycles showed that the sensitivity of detection was enhanced on the cost of specificity by increasing the number of PCR cycles. The highest amplification yield of specific fragments, which could be easily distinguished from non-specific fragments of lower intensity on agarose gels, was obtained using 35 cycles. The optimal concentration of T-DNA and gene-specific primers was tested in a range from 0.25 to $2.5\text{ } \mu\text{M}$, as shown for detection of the *At1g75960* insertion mutation in Figure 2(b). The amplification of non-specific fragments was reduced to a minimum using $0.5\text{ } \mu\text{M}$ gene-specific primer in combination with $0.25\text{ } \mu\text{M}$ FISH1 (or FISH2) T-DNA end primer in the PCR reactions.

To ascertain the sensitivity limitation of the PCR assay, a series of template dilutions ranging from 1:1000 to 1:7000 was prepared by mixing DNA from the *pad1* insertion mutant line with wild-type Col-0 DNA. Figure 2(c) illustrates that specific amplification of a T-DNA tagged *PAD1* gene fragment was observed even at 1:7000 dilution using $1\text{--}4\text{ ng } \mu\text{l}^{-1}$ template DNA. This indicated that the limiting concentration of mutant template DNA was below $0.14\text{ pg } \mu\text{l}^{-1}$, but the highest yield was achieved when the PCR was performed with $4\text{ ng } \mu\text{l}^{-1}$ template, which carried the mutant DNA diluted with wild-type DNA to 1:4000 or 1:5000. The PCR reactions for long-range amplification with LA-Taq were standardized for 35 cycles, each including 30 sec at 95°C followed by 8.5 min at 68°C . These adjustments allowed a highly reproducible recovery of T-DNA insert junctions from a very low amount of template DNA and subsequent detection of PCR products on agarose gels

Table 1 Segregation analysis of the T-DNA encoded hygromycin-resistance marker in 100 randomly chosen M_2 families

Hyg ^R :Hyg ^S *	Number of lines
3:1	64
15:1	16
63:1	2
3:1 to 15:1	7
2:1	5
<2:1	6

* χ^2 analysis for each class indicated no significant deviation from the expected ratio at a level of $P=0.05$. At least 500 M_2 seedlings were scored in each family.

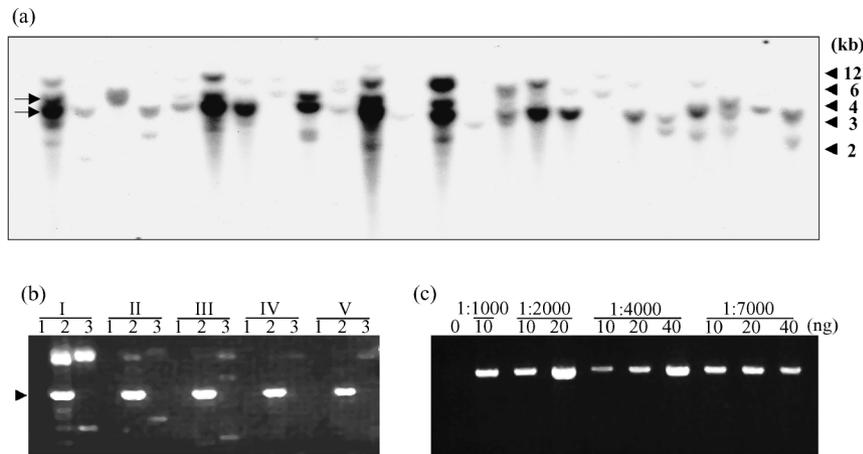


Figure 2. Estimation of average T-DNA copy number and optimization of PCR reactions.

(a) Autoradiography of a representative Southern hybridization of *Eco*RI-digested total DNAs from 24 M_1 lines with the *Xba*I–*Bam*HI right border fragment of pPCV6NFHyg T-DNA (Koncz *et al.*, 1989). Bands marked by arrows correspond to junction fragments of 5.4 kb carrying the left and right borders of direct T-DNA repeats, whereas fragments of 4.2 kb represent junctions between two right borders of inverted T-DNA repeats.

(b) The optimization of PCR primer concentrations was performed with a P1000 pool containing a single line with an insertion mutation in the gene *At1g75960*. The PCR reactions included either the gene-specific primer FSKP1 alone (1), or the T-DNA end primer FISH2 alone (3), or both (2). Different FSKP1/FISH2 primer ratios were tested in amplifications (I) 2.5/2.5 μ M; (II) 2.5/0.5 μ M; (III) 0.5/0.5 μ M; (IV) 0.5/0.25 μ M; (V) 0.25/0.25 μ M. An arrow labels the gene-specific DNA fragment. Note an absence of amplification of smaller unspecific fragments in primer combinations IV and V.

(c) Testing the detection limit of a T-DNA insertion in a known gene. DNA from the *pad1* insertion mutant line was diluted with wild-type DNA to 1000-, 2000-, 4000- and 7000-fold, and PCR reactions were performed using either 10, 20 or 40 ng DNA template in 10 μ l reaction volume with the gene-specific primer FASP2 and T-DNA end primer FISH1. Note that a significant enrichment of gene-specific amplified fragment is still observed with 40 ng template at 1 : 4000 dilution of the mutant DNA, whereas this fragment, although detected, is not enriched at 1 : 7000-fold dilution.

by ethidium bromide staining. Table 2 summarizes the advantages of this protocol in comparison to other PCR-based mutant screening techniques.

Primer design

To design gene-specific PCR primers, the following criteria were used.

(i) Optimally, the primers extended in length from 25 to 30 nucleotides and included 13 G or C.

(ii) The 5' gene-specific primer was placed 1 nt to a maximum of 300 nt upstream of the ATG codon, whereas the 3' primer was matched with the position of the stop codon.

(iii) Primers that could form self-annealing dimers, hairpins or heterodimers with the T-DNA end primers were identified (e.g. with the DNASTAR PRIMER SELECT program) and discarded. To compensate for potential incompatibility between the T-DNA end primers FISH1 or FISH2 and some gene-specific primers, several

Table 2 Comparison of PCR parameters employed by different reverse-genetics screening procedures

PCR screening methods	McKinney <i>et al.</i> (1995)	Krysan <i>et al.</i> (1996)	Winkler <i>et al.</i> (1998)	Krysan <i>et al.</i> (1999)	Present work
DNA extraction procedure	CTAB	SDS	CTAB		SDS/CsCl
Lines in the higher pool	100	1300	6000*	2025	4000/5000
Copies of T-DNA in one PCR reaction	≈2000	≈120	70–200*	≈240	≈40
Primer concentration (μ M)	0.5	0.24	0.05	0.24	Gene primer 0.5 T-DNA primer 0.25
Annealing temperature ($^{\circ}$ C)	42	65	60	65	68
Enzyme	Taq (Promega)	X-Taq (Pan Vera)	Taq (BRL)	X-Taq (Pan Vera)	LA Taq (Takara)
Detection procedure	Southern	Southern	Dot blot Digoxigenin	Southern	Ethidium bromide
Transformed lines	5300	9100	6000	60 480	90 000
Hits/assayed genes	2/10	17/63	12/70		87/154

Note differences in sensitivity of PCR reactions, stringency of primer annealing and costs of detection.

*A test PCR is performed to assay for the presence of an insert in the entire collection.

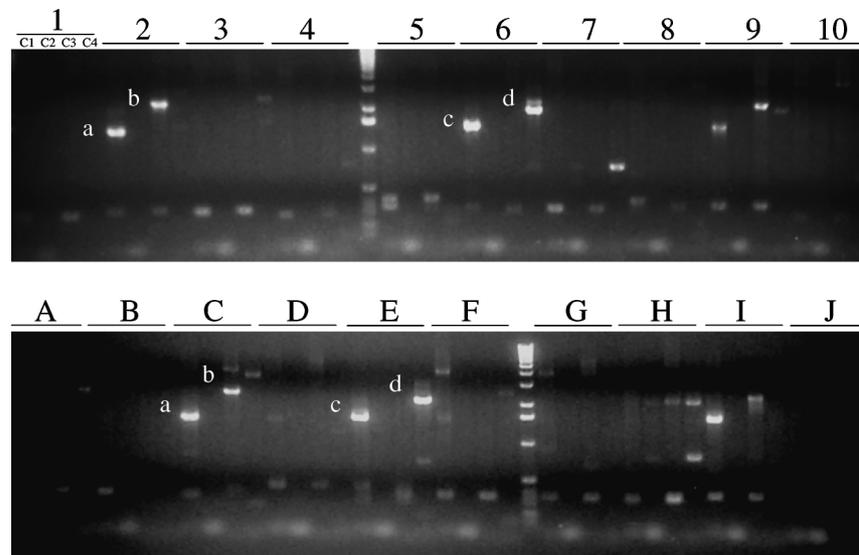


Figure 3. Identification of super-pools carrying a T-DNA insertion in the target gene.

PCR products generated with the number- and letter-coded super-pools were size separated on agarose gel. Each super-pool is represented by four samples loaded in the same order, corresponding to PCR reactions C1, C2, C3 and C4 that were performed with four different combinations of gene-specific and T-DNA end primers. It was first inspected whether DNA fragments showing higher intensity than the non-specific amplification products (appearing as weaker 'background' bands) occur repeatedly in a letter- and a number-encoded super-pool. Subsequently, it was tested whether the additive size of these putative specific fragments corresponds to the known distance between the gene-specific 5' and 3' primers. Fragments marked by letters in pools 2 and C, as well as in pools 6 and E, satisfied these conditions. (Note that pools 9 and I also contain a pair of DNA fragments corresponding to a third insertion allele, which we did not mark to prompt their identification by the reader.)

alternative primers were designed for the left and right T-DNA borders (see Experimental procedures).

- (iv) Each gene specific primer was tested in a BLASTN search with the *Arabidopsis* DNA database to eliminate those that showed an identity of 10 or more nucleotides at their 3' end with a heterologous genomic sequence.

Preparation of template DNA array and standardization of rapid screening procedure

The P100 DNA pools, each representing 100 mutant lines, were arranged in nine arrays of 10 × 10 format. P100 samples from each row (A–J) and column (1–10) of four and five arrays, respectively, were pooled to create two series of super-pools, P4000 and P5000, which contained DNA from 4000 or 5000 plants in a single PCR template. Thus each P100 sample was included twice in the letter- and number-coded super-pools, and the intersections between rows and columns corresponded to either four or five samples of P100 pools (Figure 1). Each PCR reaction contained 40 ng DNA, providing about 40 copies of any T-DNA tagged locus. The entire collection of 90 000 plants was screened with 2 × 20 × 4 PCR reactions, if all four combinations of T-DNA and gene-specific primers (gene 5' with FISH1 and FISH2; gene 3' with FISH1 and FISH2 primers, Figure 1) were used with the P4000 and P5000 templates by resolving the amplified DNA fragments on two agarose gels.

If a single-copy T-DNA tag was ideally present in the target gene, an amplification of two DNA fragments was observed with the left and right T-DNA end primers in two out of four PCR reactions with a super-pool. Some insertions containing tandem T-DNA copies resulted in the amplification of two DNA fragments with either the left or right T-DNA border primer. Rarely, two specific fragments were detected, but their additive size did not confirm the expected size of the target gene, as one of the fragments carried aberrantly transferred sequences from the vector backbone of pPCV6NFHyg. Finally, one amplified fragment was detected in those cases when only one T-DNA end was located in the target gene. As each super-pool was assayed twice, the amplified fragments detected in a letter-coded pool appeared repeatedly in a number-coded pool. To evaluate the results before confirmatory DNA sequencing, the following practical rules were applied: (i) only strong bands clearly differing from the background were considered; (ii) amplification of DNA fragments of identical size was expected in a letter- and a number-coded super-pool with the same primer combination; and (iii) insertion in a gene was ideally expected to yield two DNA fragments, indicating the position of the tag within the gene. Amplification of a DNA fragment of 7.0 kB was an indicative for trimeric tandem insertions in → (or reverse) configuration, which rarely occur in the collection and do not disturb the analysis, as usually the expected PCR products range in size

between 0.5 and 6 kb. Figure 3 shows an example for screening of the P4000 super-pools in which the DNA bands in pools 2 and C, as well as those in pools 6 and E, fulfilled these expectations.

As the intersections between letter- and number-coded super-pools corresponded to either four or five P100 pools, a second round of PCR screening was performed to identify a single P100 pool carrying the tagged gene. The second screen was carried out with $0.2\text{--}0.3\text{ ng }\mu\text{l}^{-1}$ of P100 DNA templates and produced usually two- to fivefold the amount of products observed in the first PCR round. An aliquot from the second PCR reaction was used for reamplification of DNA fragments from the positive P100 pool using 1 : 1000 dilution of the template DNA. The obtained DNA fragments were isolated to sequence the insert junctions with T-DNA specific primers. To identify a single M_2 family carrying the sequenced T-DNA tag, seeds collected from each individual M_1 plant comprising the P100 pool were germinated on hygromycin-containing medium for 15 days (Koncz *et al.*, 1994). The samples were arranged in a 10×10 array, and seedlings were collected in letter- and number-coded P10 pools to prepare 20 DNA templates (Rogers and Bendich, 1985). As described above, the row/column intersection identified the segregating M_2 family that carried the sequenced insertion mutation. Individual hygromycin-resistant plants from this M_2 line were planted in soil for further genetic analysis and screening for homozygous mutant lines. The time chart of screening protocol is depicted in Table 3.

Screening for mutations in known genes

To determine the mutation recovery rate, 39 700 plants were screened by different laboratories (see Acknowledgements) for T-DNA insertions in 154 different genes with predicted functions in regulation of protein transport to chloroplast, cell cycle, ubiquitin-dependent protein degradation, carbon catabolic repression, transcription and other cellular processes (see Experimental procedures). These mutant screens identified 87 independent insertions in 73 genes, which were characterized by sequencing of at least one of the T-DNA insert junctions. The analysis of insert

distribution in tagged genes showed that 54% of T-DNA inserts landed in exons, 23% in introns, and 23% in promoter or 5' untranslated mRNA coding sequences. Compilation of the sequenced insert junctions revealed that only 8% of insertions resulted in the amplification of two DNA bands with the left and right T-DNA border primers. By contrast, two left-border sequences of tandem T-DNA copies were found in 32% of tagged genes, whereas two right border-gene junctions were detected in 5% of mutant loci. Other tagged genes were detected only by one amplified DNA fragment, which carried either a left (41%) or a right (14%) T-DNA border junction. In few loci, where the T-DNA insertion was found within the gene, a lack of amplification with one of the T-DNA end primers indicated that deletions removed either the FISH1 or the FISH2 primer site from the left or right T-DNA ends, respectively. As suggested by the discrepancy between the estimated number of insertion loci and T-DNA copies, these data also indicated that a significant proportion of insertion loci contained tandem T-DNA copies. As observed in other T-DNA insertion mutant collections (Castle *et al.*, 1993; Krysan *et al.*, 1996; Krysan *et al.*, 1999), the tandem T-DNA copies formed predominantly inverted repeats facing the insert junctions with their left border sequences.

Estimation of the mutation rate

To calculate the probability of finding a T-DNA tag in a gene of given size, Krysan *et al.* (1999) used the equation:

$$P = 1 - [1 - (x/125\,000)]^n \quad \text{eqn 1}$$

where P is the probability of at least one insert in a gene of x kb; n is the number of inserts in the T-DNA tagged collection; and 125 000 is the haploid genome size. This formula is based on the assumption that the T-DNA integration is random and is thus similar to

$$P = 1 - f^n \quad \text{eqn 2}$$

where f is the complement of mutation frequency and n is the number of families required to identify at least one mutation (Rédei and Koncz, 1992). Whereas equation 2 ignores the size of target genes, equation 1 does not take account of the mutation rate, but assumes that mutations occur with identical frequency in any target sequence of identical size anywhere in the genome.

As the estimated average number of insertion loci was 1.29 per plant, our test population of 39 700 plants was calculated to carry $n = 51\,213$ insertions. The probability values for finding a mutation within genome segments of different size were computed using equation 1 and compared to the frequency of T-DNA tags observed in genes of different size in our PCR screens (Table 4). Except for genes smaller than 1 kb, the size distribution of genes tested in our PCR screens was comparable to that of 25 554

Table 3 Time chart of the PCR mutant screening protocol

Day	Process
1	First PCR round with P4000 and P5000 super-pools
2	Second PCR round with P100 pools
3–7	Preparative PCR and sequencing the T-DNA insert junctions
8–23	Germination and growth of individual lines from the identified P100 pool
24–25	DNA extraction from 20 P10 pools, PCR screening for a single M_2 line carrying the sequenced T-DNA tag

Table 4 Comparison of frequency of T-DNA tagged genes observed in PCR screening experiments with those expected by application of eqn 1 [$P = 1 - (1 - x/125\,000)^n$], assuming random T-DNA integration

Gene size class (kb)	<i>Arabidopsis</i> genes in given class			Genes tested			Expected frequency of random mutations in genes tested (%)	Observed frequency (%)	
	Number	Average size (kb)	Frequency (%)	Number	Average size (kb)	Frequency (%)		Mutations in genes	Genes with promoters
<1	6028	0.634	23.6	17	0.770	11.0	27.1	23.5	41.2
1–2	9627	1.479	37.7	58	1.482	37.7	45.5	32.8	50.0
2–3	5489	2.426	21.5	45	2.454	29.3	63.4	35.6	42.2
3–4	2192	3.441	8.5	17	3.491	11.0	76.1	64.7	64.7
>4	2218	5.612	8.7	17	4.822	11.0	86.1	47.1	52.9

Size distribution and frequency of genes screened for mutation were compared to those of re-annotated *Arabidopsis* genes displayed in the GenBank.

unique *Arabidopsis* genes displayed in the reannotated GenBank genome sequence database. To compare our results with the GenBank data on gene size distribution, first we accounted only those mutations that were found between ATG and stop codons in genes. Except for genes larger than 4 kb, the observed probability values for finding a T-DNA tag in a gene of a given size class showed a good correlation with the expected probability values calculated using equation 1. However, when we also accounted for those T-DNA tags which were located within 300 bp 5' upstream of ATG codons in the promoter regions of genes, closely similar frequencies were obtained for all size categories, except for genes of 3–4 kb. Whereas the frequency of all T-DNA tags recovered between ATG and stop codons was 37.7%, the additive frequency of tags in genes and their promoters was as high as 47.4%.

In conclusion, we observed 87 mutations in 154 genes, but recovered only 73 tagged unique genes because multiple insertions occurred in 13 genes. Hence we observed at least one mutation in a fraction of $P = 73/154 = 0.474$ of genes using $n = 51\,213$ T-DNA insertions in our test population. By sequencing the insert junctions of 1000 random T-DNA tags Szabados *et al.* (2002) have estimated that a $t = 0.478$ fraction of T-DNA tags results in knockout mutations. Using this t -value as an estimate for the mutation rate ($t/25.554$) in equation 2, the resulting equation:

$$P = 1 - (1 - (t/25\,554))^n \quad \text{eqn 3}$$

suggests that one would find at least one mutation in a T-DNA tagged population carrying $n = 51\,213$ inserts at a probability of 61.6%. Thus, according to equation 3, by testing 154 genes we could have expected to recover 95 independent gene mutations (about 14% more than actually detected). In fact, we could have missed some promoter mutations in our pilot experiment, as the majority of 5' gene-specific primers were designed to match with the ATG codons. Thus insertions in promoters were only detected by PCR amplification of larger DNA fragments with the

T-DNA and 3' gene-specific primers. In addition, a possible failure of detecting T-DNA tags with deleted ends and improper primer design in some cases could reduce mutation recovery. Therefore, solving equation 3 with $P = 73/154$ for the t -value results in a mutation rate estimate of 0.32, which is smaller than that observed by Szabados *et al.* (2002). Using this reduced t -value in equation 3 suggests that our collection of 90 000 plants, carrying an estimated $n = 116\,100$ T-DNA insertions, would be sufficient to detect at least one mutation in 76.5% of *Arabidopsis* genes. To reach a probability (P) of 95%, one would require about 239 142 T-DNA insertions. Nonetheless, it remains to be assessed whether this is a precise estimate by further exploitation of the described rapid PCR mutant screening technique.

Experimental procedures

Growth and transformation of *Arabidopsis*

Inflorescences of *Arabidopsis thaliana* (Col-0) plants were transformed using vacuum infiltration (Bechtold *et al.*, 1993) with *Agrobacterium* GV3101 (pMP90RK) carrying the binary vector pPCV6NFHyg (Koncz and Schell, 1986; Koncz *et al.*, 1989). This vector was chosen for insertion mutagenesis, as the presence of a promoterless kanamycin resistance gene at the right T-DNA border also provided a possibility to select for transformants that express in-frame translational NPT-II fusions as a result of T-DNA integration into coding domains of plant genes (Koncz *et al.*, 1989). The infiltration medium [half concentration of Murashige–Skoog (MS) basal salts with B5 vitamin; Sigma, pH 5.7) (Sigma-Aldrich Co., Taufkirchen, Germany), 5% sucrose, 0.05 μM benzylaminopurine, and 0.005% Silwet L-77) was supplemented with 100 mg l^{-1} carbenicillin in order to select for the bacterial antibiotic resistance marker of pPCV6NFHyg, and thereby for the maintenance of *Agrobacterium* binary vector *in planta*. This modification increased the transformation frequencies from 0.5 ± 0.4 to $3 \pm 1.2\%$ (data not shown). The M_1 seed progeny of 11 250 transformed plants were pooled. Seed aliquots were surface sterilized with 5% calcium hypochloride solution containing 1% Triton-X 100 and germinated on seed germination medium (Koncz *et al.*, 1994) containing

15 mg l⁻¹ hygromycin (Roche Diagnostics GmbH, Mannheim, Germany) to select for transformed seedlings. The seedlings were grown in glass jars for 2 weeks in a growth chamber at 23°C under 200 µEinstein m⁻² sec⁻¹ irradiance using an 8 h light, 16 h dark cycle, then planted into soil and grown for additional 2 weeks under short-day conditions to obtain large leaf rosettes. Upon shifting to long-day (16 h light/8 h dark) conditions to induce flowering, 200 mg leaf material was collected from each plant bearing fruits to obtain pooled material from every 100 plants. To determine the segregation ratio of the T-DNA-encoded hygromycin-resistance gene, seedlings from 100 randomly selected M₂ families were germinated as described above. DNAs prepared from hygromycin-resistant seedlings of 24 M₂ families were used to estimate the T-DNA copy number using Southern hybridization, as described (Koncz *et al.*, 1989).

Preparation of pooled DNA arrays

From 900 P100 pools of leaf material, each containing samples from 100 individual plants, total DNA was extracted (Dellaporta *et al.*, 1983) and purified by CsCl ethidium bromide gradient centrifugation (Sambrook *et al.*, 1989). The P100 samples were arranged in nine 10 × 10 quadrats, each carrying 100 P100 DNA pools representing 10 000 plants (Figure 1). The collection was divided into two portions corresponding to four and five quadratic arrays, then samples containing an identical amount of DNA were pooled from each number-coded row and letter-coded column of the arrays to create 20 P4000 and 20 P5000 super-pools, respectively. Once a P100 pool carrying a gene mutation was identified, DNA samples from seedlings representing 100 individual M₂ families in the P100 pool were isolated using a cetyltrimethylammonium bromide (CTAB) precipitation protocol (Rogers and Bendich, 1985). The DNA samples were again arranged in a 10 × 10 array to create 20 P10 pools by combining samples from rows and columns.

PCR primers

The PCR screens were performed with the T-DNA left border (FISH1, 5'-CTGGGAATGCGCAAATCAAGGCATC-3') and right border (FISH2, 5'-CAGTCATAGCCGAATAGCCTCTCCA-3') primers. When these primers were complementary to some gene-specific primers, alternative left border (HOOK1, 5'-CTACACTGAATTGGTAGCTCAAAGTGC-3' or HOOK3, 5'-GTTGACAGACTGCCTAGC-ATTTGAGTGTG-3') and right border (HOOK2, 5'-TACTTTCTCGGCA-GGAGCAAGGTGA-3' or HOOK4 5'-TCAGAGCAGCCGATTGTCT-GTTGTG-3') primers were used. For optimization of PCR conditions, specific amplification of T-DNA junctions from the *prl1* mutant DNA (Németh *et al.*, 1998) was assayed with the PRLX (5'-ATCTGTCCGAGCAATGACCCTCCAT-3') and PRLY (5'-GGG-AAGAGCACCACATTATACCTG-3') primers. Similarly, primers FISH2 and FSKP1 (5'-CGTTTCCATAACCATGTCTGCGAAGA AG-3') were used to detect amplification of a 1.2 kb T-DNA junction fragment of tagged At1g75960 gene, whereas a T-DNA junction fragment of 2 kb from the mutant *pad1* gene was amplified with the FISH1 and FASP2 (5'-GTTTCTTCGACAGGGCCTTTCTTGG-3') primers.

PCR mutant screening protocol

The initial optimization of PCR conditions was performed using DNA from the *prl1*, At1g75960 and *pad1* T-DNA mutants in combination with TaqI (Roche), Elongase (Invitrogen GmbH, Karlsruhe,

Germany) and LA-Taq (Takara Shuzo Co. represented by BioWhittaker Europe, Taufkirchen, Germany) DNA polymerases. As the highest sensitivity and yield was obtained with LA-Taq independently of the size of expected PCR products, further standardization of the PCR reactions was performed with this enzyme by testing different annealing conditions (from 55 to 68°C), primer concentrations (from 0.25 to 2.5 µM) and template DNA dilutions (from 1:1000 to 1:7000), as described in the text. The standardized PCR assays for mutant screening were performed with Takara buffer containing 2 mM MgCl₂, 0.2 mM dNTP, 0.5 µM gene-specific primer, 0.25 µM T-DNA end primer and 4 ng µl⁻¹ DNA template. The PCR reactions were initiated by heating the samples for 5 min at 95°C, followed by 35 cycles of amplification for 30 sec at 95°C and 8.5 min at 68°C, and terminated by a final elongation step for 10 min at 68°C. The samples were separated on 1.2% agarose gels containing ethidium bromide, and images of DNA fragment patterns were recorded using a DC120 digital camera and Kodak DIGITAL SCIENCE software.

The presence of a T-DNA tag in a target gene was indicated by the amplification of DNA fragments of identical size in two samples representing a number- and a letter-coded P4000 or P5000 super-pool. As the identified super-pool corresponded to either four or five P100 pools, a second round of PCR was performed with 2.5 ng P100 DNA in 20 µl PCR reaction mix, as described above. The PCR reaction mixture of P100 pool producing a reamplification of expected DNA fragments was diluted 1000-fold to perform a preparative PCR using 0.5 µM of both gene-specific and T-DNA end primers. Subsequently, the DNA fragments were isolated and sequenced with the T-DNA end primers. To identify a single mutant line carrying the mutation in a P100 pool, at least five hygromycin-resistant 15-day-old-seedlings were collected from each individual M₂ family for preparation of 20 P10 DNA templates using the pooling strategy described above. The CTAB-purified DNA pools were dissolved in 100 µl buffer (1 mM Tris-HCl pH 8.0, 0.1 mM EDTA), and 1 µl from each DNA sample was used as template in a PCR reaction of 10 µl containing 0.5 µM gene-specific and T-DNA end primers, as described above. Hygromycin-resistant seedlings from the M₂ family segregating the identified mutation were planted into soil for harvesting leaf material from at least 15 individual plants. Screening for lines carrying the T-DNA induced mutation in homozygous form was performed with CTAB-purified DNA samples (Rogers and Bendich, 1985) using either the gene-specific primers or their combinations with T-DNA specific primers. Putative homozygous lines (producing the expected T-DNA junction fragments, but no PCR product with gene specific primers in PCR reactions performed with Taq polymerase) were subjected to further analysis by Southern DNA hybridization.

Estimation of mutation recovery rate

To estimate the mutation frequency, the collection was screened for T-DNA insertions in 154 genes. At least one mutation was detected in genes At1g01720, At1g02970, At1g02980, At1g04940, At1g10210, At1g10650, At1g10970, At1g11890, At1g13440, At1g15730, At1g26830, At1g44170, At1g49620, At1g51660, At1g70210, At1g71020, At1g73690, At2g01500, At2g04660, At2g19190, At2g21950, At2g24820, At2g25490, At2g29960, At2g33610, At2g37340, At2g41070, At2g47840, At3g10540, At3g12280, At3g13730, At3g16620, At3g16650, At3g17950, At3g25250, At3g43210, At3g46740, At3g50070, At3g50530, At3g50660, At3g51260, At4g02570, At4g03270, At4g08320, At4g11150, At4g14970, At4g19690, At4g24280, At4g26070, At4g28480, At4g28980, At4g34240, At4g36380, At4g37630,

At5g02110, At5g04510, At5g09890, At5g10440, At5g16620, At5g19610, At5g20010, At5g25350, At5g27620, At5g39440, At5g39500, At5g45260, At5g46010, At5g48000, At5g49910, At5g55710, At5g62540, At5g65420, At5g66840; whereas no mutation was found in genes At1g06940, At1g09020, At1g10470, At1g14400, At1g16890, At1g20700, At1g20710, At1g20930, At1g22640, At1g23080, At1g35860, At1g43140, At1g59580, At1g66750, At1g69670, At1g70940, At1g76540, At1g77000, At1g77110, At1g78870, At2g01110, At2g01420, At2g02760, At2g16640, At2g22490, At2g28610, At2g32850, At2g33880, At2g42880, At2g44060, At2g47620, At3g01090, At3g03660, At3g10220, At3g11260, At3g13550, At3g17590, At3g18010, At3g18040, At3g21220, At3g23710, At3g29160, At3g30180, At3g42830, At3g63130, At4g02510, At4g03320, At4g11260, At4g13980, At4g15900, At4g22540, At4g23570, At4g26020, At4g29510, At4g29810, At4g33350, At4g34160, At4g35550, At5g02300, At5g05000, At5g05770, At5g08130, At5g09420, At5g15100, At5g16530, At5g17810, At5g18590, At5g19320, At5g20300, At5g20570, At5g23260, At5g38970, At5g40440, At5g42390, At5g45980, At5g46210, At5g50920, At5g52440, At5g55190, At5g55910, At5g67260. The size distribution of *Arabidopsis* genes was calculated by examining all unique predicted gene sequences available in the GenBank *Arabidopsis* database.

Public access to the collection

The collection is made freely available for visitors to perform mutant screens requiring a local purification of P100 pools at the Max-Planck Institut für Züchtungsforschung. PCR templates from super-pools and necessary instructions are mailed for primary screens. Seeds from P100 pools are available as individual M_2 families.

Acknowledgements

We thank Drs A. Bachmair, S. Bancos, D. Bartels, L. Bögre, L. Deslandes, P. Dudley, P. Genschik, T. Hilbricht, H. Hirt, J. Huang, R. P. Jarvis, G. Jürgens, J. Kellmann, H.-H. Kirch, J. Martinsson, J. Marques, T. Mészáros, R. Müller, K. Nettesheim, O. Olsson, M. Paape, K. Palme, Y. Parmentier, A. Páy, R. Rigó, E. Scheickl, R. Simon, I. Somssich, K. Steinborn, M. Szekeres, J. Uhrig, M. Umeda, M. Yamaguchi, A. Zilberstein, and I. Zimmermann for providing data on mutations they identified in our collection. This work was supported by grants from the European Commission (QLRT-2000-01871), Deutsche Forschungsgemeinschaft (KO1483/3-1) and Human Frontiers Science Program (grant no. RG00162-2000) for C. Koncz, and a Marie Curie Fellowship (EU HPMF-CT-2000-00597) for G. Ríos.

References

Aarts, M.G., Corzaan, P., Stekema, W.J. and Pereira, A. (1995) A two-element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **247**, 555–564.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815(2001).

Azpiroz-Leehan, R. and Feldmann, K. (1997) T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. *Trends Genet.* **13**, 152–156.

Balzegue, S., Dubreucq, B., Chauvin, S. et al. (2001) Improved PCR-walking for large-scale isolation of plant T-DNA borders. *Biotechniques*, **30**, 496–504.

Bancroft, I. and Dean, C. (1993) Transposition pattern of the maize element *Ds* in *Arabidopsis thaliana*. *Genetics*, **134**, 1221–1229.

Bancroft, I., Bhatt, A.M., Sjodin, C., Scofield, S., Jones, J.D.G. and Dean, C. (1992) Development of an efficient two-element transposon tagging system in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **233**, 449–461.

Baumann, E., Lewald, J., Saedler, H., Schulz, B. and Wisman, E. (1998) Successful PCR-based reverse genetic screens using an *En-1* mutagenized *Arabidopsis thaliana* population generated via single-seed descent. *Theor. Appl. Genet.* **97**, 729–734.

Bechtold, N., Ellis, J. and Pelletier, G. (1993) In planta *Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *Comptes Rend. L'Acad. Sci. Serie III Sci. Vie*, **316**, 1194–1199.

Bouché, N. and Bouchez, D. (2001) *Arabidopsis* gene knockout: phenotypes wanted. *Curr. Opin. Plant Biol.* **4**, 111–117.

Bouchez, D. and Höfte, H. (1998) Functional genomics in plants. *Plant Physiol.* **118**, 725–732.

Castle, L.A., Errampalli, D., Atherton, T.L., Franzmann, L.H., Yoon, E.S. and Meinke, D.W. (1993) Genetic and molecular characterization of embryonic mutants identified following seed transformation in *Arabidopsis*. *Mol. Gen. Genet.* **241**, 504–514.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.

Colbert, T., Till, B.J., Tompa, R., Reynolds, S., Steine, M.N., Yeung, A.T., McCallum, C.M., Comai, L. and Henikoff, S. (2001) High-throughput screening for induced point mutations. *Plant Physiol.* **126**, 480–484.

Dellaporta, S.L., Wood, J. and Hicks, J.B. (1983) A plant DNA miniprep: version II. *Plant Mol. Biol. Rep.* **1**, 19–21.

Devic, M., Hecht, V., Berger, C., Delseny, M. and Gallois, P. (1995) An assessment of promoter trapping as a tool to study plant zygotic embryogenesis. *Comptes Rend. L'Acad. Sci. Serie III Sci. Vie*, **318**, 121–128.

Farrás, R., Ferrando, A., Jásik, J., Ökrész, L., Tiburcio, A., Salchert, K., del Pozo, C., Schell, J. and Koncz, C. (2001) SKP1-SnRK protein kinase mediate proteasomal binding of a plant SCF ubiquitin ligase. *EMBO J.* **20**, 2742–2756.

Fedoroff, N.V. and Smith, D.L. (1993) A versatile system for detecting transposition in *Arabidopsis*. *Plant J.* **3**, 273–289.

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

Feldmann, K.A. and Marks, M.D. (1987) *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue approach. *Mol. Gen. Genet.* **208**, 1–9.

Forsthoefel, N.R., Wu, Y., Schulz, B., Bennett, M.J. and Feldmann, K.A. (1992) T-DNA insertion mutagenesis in *Arabidopsis*: prospects and perspectives. *Aust. J. Plant Physiol.* **19**, 353–366.

Galbiati, M., Moreno, M.A., Nadzan, G., Zourelidou, M. and Dellaporta, S.L. (2000) Large-scale T-DNA mutagenesis in *Arabidopsis* for functional genomic analysis. *Funct. Integr. Genomics*, **1**, 25–34.

Goddijn, O.J., Lindsey, K., van der Lee, F.M., Klap, J.C. and Sijmons, P.C. (1993) Differential gene expression in nematode-induced feeding structures of transgenic plants harbouring promoter-*gusA* fusion constructs. *Plant J.* **4**, 863–873.

Kiegle, E., Moore, C.A., Haseloff, J., Tester, M.A. and Knight, M.R. (2000) Cell-type-specific calcium responses to drought, salt and cold in the *Arabidopsis* root. *Plant J.* **23**, 267–278.

Koncz, C. and Schell, J. (1986) The promoter of T₁-DNA gene 5 controls the tissue specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* **204**, 383–396.

- Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kálmán, Z., Körber, H., Rédei, G.P. and Schell, J. (1989) High-frequency T-DNA-mediated gene tagging in plants. *Proc. Natl Acad. Sci. USA*, **86**, 8467–8471.
- Koncz, C., Németh, K., Rédei, G.P. and Schell, J. (1992) T-DNA insertional mutagenesis in *Arabidopsis*. *Plant Mol. Biol.* **20**, 963–976.
- Koncz, C., Martini, N., Szabados, L., Hroudá, M., Bachmair, A. and Schell, J. (1994) Specialized vectors for gene tagging and expression studies. In *Plant Molecular Biology Manual*, Vol. B2 (Gelvin, S.B., Schilperoort, R.A. and Verma, D.P.S., eds). Dordrecht, the Netherlands: Kluwer Academic Publishers, pp. 1–22.
- Krysan, P.J., Young, J.C., Tax, F. and Sussman, M.R. (1996) Identification of transferred DNA insertions within *Arabidopsis* genes involved in signal transduction and ion transport. *Proc. Natl Acad. Sci. USA*, **93**, 8145–8150.
- Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell*, **11**, 2283–2290.
- Liu, Y.G., Mitsukawa, N., Oosumi, T. and Whittier, R.F. (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457–463.
- Long, D., Martin, M., Sundberg, E., Swinburne, J., Puangsomlee, P. and Coupland, G. (1993) The maize transposable element system *Ac/Ds* as a mutagen in *Arabidopsis*: identification of an *albino* mutation induced by *Ds* insertion. *Proc. Natl Acad. Sci. USA*, **90**, 10370–10374.
- Long, D., Goodrich, J., Wilson, K., Sundberg, E., Martin, M., Puangsomlee, P. and Coupland, G. (1997) *Ds* elements on all five *Arabidopsis* chromosomes and assessment of their utility for transposon tagging. *Plant J.* **11**, 145–148.
- Mahalingam, R. and Fedoroff, N. (2001) Screening insertion libraries for mutations in many genes simultaneously using DNA microarrays. *Proc. Natl Acad. Sci. USA*, **98**, 7420–7425.
- Martienssen, R.A. (1998) Functional genomics: probing plant gene function and expression with transposons. *Proc. Natl Acad. Sci. USA*, **95**, 2021–2026.
- Mathur, J., Szabados, L., Schaefer, S., Grunenberg, B., Lossow, A., Jonas-Straube, E., Schell, J., Koncz, C. and Koncz-Kálmán, Z. (1998) Gene identification with sequenced T-DNA tags generated by transformation of *Arabidopsis* cell suspensions. *Plant J.* **13**, 707–716.
- McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000a) Targeted screening for induced mutations. *Nature Biotechnol.* **18**, 455–457.
- McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000b) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol.* **123**, 439–442.
- McKinney, E.C., Ali, N., Traut, A., Feldmann, K.A., Belostotsky, D.A., McDowell, J.M. and Meagher, R.B. (1995) Sequence-based identification of T-DNA insertion mutations in *Arabidopsis*: actin mutants *act2-1* and *act4-1*. *Plant J.* **8**, 613–622.
- Meissner, R.C., Jin, H., Cominelli, E. et al. (1999) Function search in a large transcription factor gene family in *Arabidopsis*: assessing the potential of reverse genetics to identify insertional mutations in R2R3 MYB genes. *Plant J.* **11**, 1827–1840.
- Németh, K., Salchert, K., Putnoky, P. et al. (1998) Pleiotropic control of glucose and hormone responses by *PRL1*, a nuclear WD protein in *Arabidopsis*. *Genes Devel.* **12**, 3059–3073.
- Okamoto, H. and Hirochika, H. (2000) Efficient insertion mutagenesis of *Arabidopsis* by tissue culture-induced activation of the tobacco retrotransposon *Tto1*. *Plant J.* **23**, 291–304.
- Osborne, B.I., Wirtz, U. and Baker, B. (1995) A system for insertional mutagenesis and chromosomal rearrangement using the *Ds* transposon and *Cre-lox*. *Plant J.* **7**, 687–701.
- Parinov, S. and Sundareshan, V. (2001) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the sequencing project. *Curr. Opin. Biotechnol.* **11**, 157–161.
- Parinov, S., De Sevugan, M.Y., Yang, W.C., Kumaran, M. and Sundareshan, V. (1999) Analysis of flanking sequences from *Dissociation* insertion lines: a database for reverse genetics in *Arabidopsis*. *Plant Cell*, **11**, 2263–2270.
- Rédei, G.P. and Koncz, C. (1992) Classical mutagenesis. In *Methods in Arabidopsis Research* (Koncz, C., Chua, N.-H. and Schell, J., eds). Singapore: World Scientific, pp. 16–82.
- Rogers, S.O. and Bendich, A.J. (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol. Biol.* **5**, 69–76.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., Caboche, M. and Lecharny, A. (2002) FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucl. Acids Res.* **30**, 94–97.
- Smith, D., Liu, Y.-G., Yanai, Y., Ishiguro, S., Okada, K., Shibata, D., Whittier, R.F. and Fedoroff, N.V. (1996) Characterization and mapping of *Ds*-GUS-T-DNA lines for targeted insertional mutagenesis. *Plant J.* **10**, 721–732.
- Speulman, E., Metz, P.L., van Arkel, G., te Lintel Hekkert, B., Stiekema, W.J. and Pereira, A. (1999) A two-component *enhancer-inhibitor* transposon mutagenesis system for functional analysis of the *Arabidopsis* genome. *Plant Cell*, **11**, 1853–1866.
- Speulman, E., van Asperen, R., van der Laak, J., Stiekema, W.J. and Pereira, A. (2000) Target selected insertional mutagenesis on chromosome IV of *Arabidopsis* using the *En-1* transposon system. *J. Biotechnol.* **78**, 301–312.
- Sundareshan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D., Dean, C., Ma, H. and Martienssen, R. (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Devel.* **9**, 1797–1810.
- Szabados, L., Kovács, I., Oberschall, A. et al. (2002) Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. *Plant J.* **32**, 233–242.
- Tax, F.E. and Vernon, D.M. (2001) T-DNA associated duplication/translocations in *Arabidopsis*. Implications for mutant analysis and functional genomics. *Plant Physiol.* **126**, 1527–1538.
- Thornycroft, D., Sherson, S.M. and Smith, S.M. (2001) Using gene knockouts to investigate plant metabolism. *J. Exp. Bot.* **52**, 1593–1601.
- Tissier, A.F., Merillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G. and Jones, J.D.G. (1999) Multiple independent defective *Suppressor-mutator* transposon insertions in *Arabidopsis*: a tool for functional genomics. *Plant Cell*, **11**, 1841–1852.
- Topping, J.F., Wei, W. and Linsey, K. (1991) Functional tagging of regulatory elements in the plant genome. *Development*, **112**, 1009–1019.
- Weigel, D., Ahn, J.H., Blázquez, M.A. et al. (2000) Activation tagging in *Arabidopsis*. *Plant Physiol.* **122**, 1003–1013.
- Winkler, R.G., Frank, M.R., Galbraith, D.W., Feyereisen, R. and Feldmann, K.A. (1998) Systematic reverse genetics of transfer-DNA-tagged lines of *Arabidopsis*. *Plant Physiol.* **118**, 743–750.

- Wisman, E., Cardon, G.H., Fransz, P. and Saedler, H.** (1998a) The behaviour of the autonomous maize transposable element *En/Spm* in *Arabidopsis thaliana* allows efficient mutagenesis. *Plant Mol. Biol.* **37**, 989–999.
- Wisman, E., Hartmann, U., Sagasser, M., Baumann, E., Palme, K., Hahlbrock, K., Saedler, H. and Weisshaar, B.** (1998b) Knock-out mutants from an *En-1* mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proc. Natl Acad. Sci. USA*, **95**, 12432–12437.
- Yephremov, A. and Saedler, H.** (2000) Display and isolation of transposon flanking sequences starting from genomic DNA or RNA. *Plant J.* **21**, 495–505.
- Young, J.C., Krysan, P.J. and Sussman, M.R.** (2001) Efficient screening of *Arabidopsis* T-DNA insertion lines using degenerate primers. *Plant Physiol.* **125**, 513–518.