

TECHNIQUES FOR MOLECULAR ANALYSIS

# Establishing gene function by mutagenesis in *Arabidopsis thaliana*

Lars Østergaard\* and Martin F. Yanofsky

Section of Cell and Developmental Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA

Received 1 December 2003; revised 23 February 2004; accepted 4 March 2004.

\*For correspondence (fax +1 858 822 1772; e-mail [larsoe@biomail.ucsd.edu](mailto:larsoe@biomail.ucsd.edu)).

## Summary

The nuclear genome of *Arabidopsis thaliana* was sequenced to near completion a few years ago, and ahead lies the challenge of understanding its meaning and discerning its potential. How many genes are there? What are they? What do they do? Computer algorithms combined with genome array technologies have proven efficient in addressing the first two questions as shown in a recent report (Yamada *et al.*, 2003). However, assessing the function of every gene in every cell will require years of careful analyses of the phenotypes caused by mutations in each gene. Current progress in generating large numbers of molecular markers and near-saturation insertion mutant collections has immensely facilitated functional genomics studies in *Arabidopsis*. In this review, we focus on how gene function can be revealed through the analysis of mutants by either forward or reverse genetics. These mutants generally fall into two distinct classes. The first class typically includes point mutations or small deletions derived from chemical or fast neutron mutagenesis whereas the second class includes insertions of transferred-DNA or transposon elements. We describe the current methods that are used to identify the gene corresponding to these mutations, which can then be used as a probe to further dissect its function.

**Keywords:** *Arabidopsis*, gene function, mutagenesis, forward genetics, reverse genetics, functional genomics.

## Introduction

In little more than a decade, research in plant biology has taken huge leaps toward understanding nearly all aspects of plant growth and development. Although significant progress has been made in many species, here we focus on *Arabidopsis thaliana*, which has been adopted as a model system for molecular-genetic studies (Meyerowitz, 1987). At the end of year 2000, an international effort to sequence the *Arabidopsis thaliana* ecotype Columbia (Col-0) genome was concluded [The *Arabidopsis* Genome Initiative (AGI), 2000] and 26 828 potential genes were predicted of which 25 540 are annotated as protein-coding (Yamada *et al.*, 2003). Based on homology searches, 30% of these predicted genes could not be assigned biological or biochemical function (AGI, 2000), and definitive functions for individual genes have been thoroughly established for less than 10%. However, with the generation of an increasing number of genetic resources in *Arabidopsis*, this percentage is bound to dramatically increase over the next few years.

In January 2000, a group of scientists met at the Salk Institute for Biological Studies, La Jolla, CA, USA to make projections and evaluations for the 'post-genomic era' (Chory *et al.*, 2000). Here, they formulated the '2010 Project,' whose goal is 'to broaden the scope to consider how genomics could be applied to understand all plants and to understand the function of all genes of a reference organism within their cellular, organismal, and evolutionary context by the year 2010' (Chory *et al.*, 2000). The ultimate goal of this effort is to be able to create a virtual plant, whose growth can be studied on a computer. Information on gene expression of any gene at any step of plant development, in any organ or cell, and in response to a series of environmental cues would be accessible (Chory *et al.*, 2000).

For the plant science community to reach the challenging goals of the '2010 project,' the study of individual genes with respect to mutant phenotypes, where and when they are expressed, how they are positioned in signal transduction

pathways, as well as biochemical and structural studies will have to be carried out. This is obviously a huge undertaking, involving plant scientists from a wide variety of research areas. In this review, we focus on how gene function can be revealed through analysis of mutants by either forward or reverse genetics, and we attempt to provide an overview of the methodologies that are currently available to obtain mutants with desired phenotypes or mutations in a gene of interest. We also describe the current methods that are used to clone the gene corresponding to a mutation and how to prove that the cloned gene is in fact the gene of interest. These important techniques allow us to investigate the genetic factors underlying the diverse processes of *Arabidopsis* biology.

### Gene function through mutagenesis

How do we determine the function of a specific gene? One way is to simply analyze its deduced protein sequence to determine if it is related to a gene of known function. For example, we may learn from such an analysis that the gene of interest encodes a putative transcription factor. While this would indeed provide valuable insight into the biochemical function of this gene product, it would provide no clues as to its specific function during plant growth and development. Expression studies, such as RNA blotting or *in situ* hybridization may provide insight into the cells and/or organs in which the gene is expected to function. Thus, if this gene is specifically expressed in flowers, for example, one could conclude that the gene functions in some aspect of flower development. However, such descriptive studies fall short of the ultimate goal of assigning a biological function to a given gene. To achieve this goal it is necessary to identify a mutation in the given gene and to compare wild-type plants with plants that harbor this mutation. For example, if the mutant plants show an alteration in flower morphology, we can conclude that the mutated gene is normally required to ensure normal flower development.

An alternative to performing mutagenesis screens is to take advantage of the natural variation that is known to occur among the hundreds of distinct ecotypes of *Arabidopsis* that have been isolated from around the world. One such example is the allelic variation among *Arabidopsis* ecotypes at the *FRI* locus that was shown to be important for flowering time control (Johanson *et al.*, 2000).

Traditional forward mutant screens were generally carried out to identify genes involved in a specific process of interest. For example, if a researcher was interested in flower development, mutant collections, generated by chemical treatment or other means, were screened for alterations in floral structure. Isolated mutants were subsequently studied and the corresponding genes were eventually cloned. More recently, reverse genetics screens are widely used as the availability of research tools has expanded. Through these

resources, it is possible to identify a gene of interest through database mining on the computer, and then to order from stock centers the corresponding mutant for phenotypic analyses. The combination of forward and reverse genetics approaches provides powerful tools that should eventually help researchers to reach the lofty goals of the '2010 project'.

### Forward genetics

The classical or 'forward genetics' way of acquiring mutants with desired phenotypes involves mutagenesis of a large number of wild-type seeds by treatment with chemical reagents or irradiation. Such treatments typically introduce single nucleotide changes or small deletions in the genome. As only two to three cells within each seed will ultimately give rise to the next generation of seeds, it is the mutations in these cells that will be carried forward into future generations.

Although it is not always practical to do so, it is always desirable to carry out a 'saturation' screen in which all possible genes that can be mutated to give a specific phenotype will be identified. The chemical mutagen, ethyl methane sulfonate (EMS), typically introduces dozens of mutations in each plant, and it is generally possible to find a mutation in any given gene by screening fewer than 5000 plants from the mutagenized M1 generation (Feldmann *et al.*, 1994; Greene *et al.*, 2003). Isolation of dominant mutations in the M1 generation has been reported (McConnell *et al.*, 2001), but this is rare. Most often recessive mutations are recovered. Generally, the M1 seeds are planted and grown to maturity and the resulting (M2) seeds are planted and screened for the desired phenotypes. A detailed description of the methods involved in performing a mutagenesis screen using EMS has previously been described (Redei and Koncz, 1992; Weigel and Glazebrook, 2002).

For example, if one is interested in identifying genes that specify petal development in flowers, one screens the M2 plants for any alterations in petal development. These are then collected for further examination in the M3 or subsequent generations. To 'clean up' the desired mutant line from the additional mutations caused by the mutating agent, it is important to backcross the mutant plants at least three times to the corresponding wild-type ecotype by following the phenotype in the F2 generations. Allelism tests are recommended with any recessive mutants of the same phenotypic class, as these putatively correspond to mutations in the same gene. In the petal example noted above, null mutations in either of the two genes, *APETALA3* (*AP3*) or *PISTILLATA* (*PI*) give rise to flowers that lack petals and stamens (Bowman *et al.*, 1989). In a screen for mutants in petal development, one would expect to obtain mutations in the *AP3* and *PI* genes as well as any additional gene required for normal petal development. By crossing two different

mutants together and analyzing the resulting F1 phenotypes it is possible to determine if the mutations are in the same or different genes. If the resulting F1 plants are of wild-type appearance then the two mutations are in different genes, whereas if F1 plants display the mutant phenotypes, one would conclude that the mutations are within the same gene. If a mutant screen is carried to saturation (see definition above) then it would be anticipated that multiple mutant alleles would be identified for each gene.

The goal of forward genetics screens is to use mutants to identify all of the genes involved in a specific process. Through the detailed phenotypic analysis of these mutants and the characterization of the corresponding genes, it is possible to begin to outline the biochemical mechanisms that underlie this process.

Even today, there are still several good reasons to screen for mutants in chemically mutagenized populations. In addition to loss-of-function mutations, chemical mutagenesis (with e.g. EMS) can result in an allelic series, allowing for strong, intermediate and weak alleles of a given gene (Bowman *et al.*, 1991). Moreover, chemically induced missense mutations can lead to temperature-sensitive phenotypes, allowing for normal growth under permissive temperatures and phenotypic analyses under restrictive temperatures (Bowman *et al.*, 1989). Furthermore, genes that produce lethal phenotypes when inactivated can be studied when less severe alleles are available. One example comes from studies of the *VTC1* (*CYT1*) gene: In one study, it was found that weak point mutant alleles of *VTC1* generated by EMS mutagenesis can result in ozone sensitivity and reduced vitamin C levels in *Arabidopsis* (Conklin *et al.*, 1999), whereas a separate study showed that *vtc1* null mutations cause embryo lethality (Lukowitz *et al.*, 2001).

Transposon insertion lines have been created (see below) and were used in forward genetics screens (Sundaresan *et al.*, 1995). In addition, many groups have used plant transformation and insertion of the transferred-DNA (T-DNA) from *Agrobacterium* to create knockout alleles. One advantage of these approaches compared with chemical mutagenesis is that the chromosomal location of the inserted DNA can be easily identified as it provides a molecular tag with known sequence. A disadvantage, however, is that often only one gene is disrupted per plant requiring a much larger population of plants to be screened and that these insertions generally lead to complete loss of gene function.

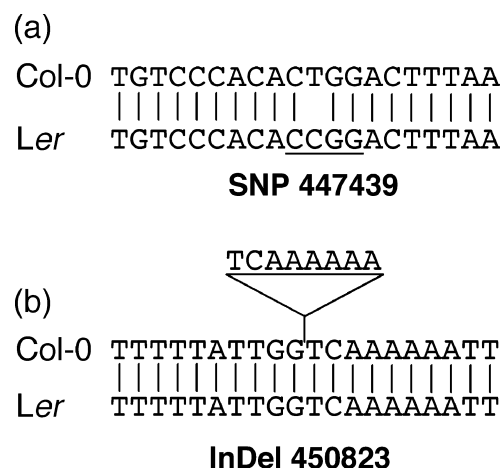
### Map-based cloning

While chemical or radiation mutagenesis offers many advantages for isolating a collection of mutants that are defective for a particular process, isolation of the corresponding genes can often be problematic because there is no molecular tag (e.g. insertion) that would allow the direct

cloning of the gene of interest. Map-based cloning, often called positional cloning, is still the most effective strategy to isolate the gene that corresponds to a chemically or radiation-derived mutant. To obtain the gene of interest, two different tools or resources are needed. The first tool involves molecular or genetic markers. These are markers whose chromosomal map positions are already known. The second tool is referred to as a mapping population that segregates for the mutation of interest. This population is most often generated by crossing the mutant plant in a specific ecotype (e.g. Columbia), to wild-type plants of a different ecotype (e.g. Landsberg *erecta*). It is generally not necessary to 'clean up' the genetic background of the mutation before crossing to the opposite genotype. The resulting F1 progeny from such a cross are grown to maturity and seeds are harvested. The recessive mutant characteristic will reappear in the F2 population, and the segregation of molecular or genetic markers in this population allows the identification of markers that are closely linked to the mutation of interest. Mapping-resolution is mainly determined by the size of the mapping population, and resolutions in the range of 10–40 kb can usually be obtained with ~1000 F2 plants (Lukowitz *et al.*, 2000). Because of the wealth of markers currently available as a public resource, analysis of the F2 plants is generally a straightforward process that requires only weeks to narrow the interval down to a handful of candidate genes. This is in stark contrast to the situation a decade ago when it typically took several years to identify the gene of interest.

Hundreds of different *Arabidopsis* ecotypes have been found around the world and are available to the plant science community through the *Arabidopsis* stock centers (<http://www.arabidopsis.org/abrc> or <http://nasc.nott.ac.uk>). Several of these accessions are sufficiently divergent to support the design of molecular markers at a high density and have been successfully used in positional cloning projects. The most commonly used combination is Landsberg *erecta* × Columbia 0 (*Ler* × Col-0). This combination was used to create a recombinant inbred map (Lister and Dean, 1993), which soon became established as the standard for genetic placement of molecular markers, and hundreds of markers have been analyzed in these lines.

Collections of markers are now available through user-friendly web interfaces. Cereon Genomics LLC (Cambridge, MA, USA) partially sequenced the Landsberg *erecta* genome by a shotgun approach (Jander *et al.*, 2002) and has generated 56 670 markers at the time of writing, corresponding to one polymorphism per approximately 2500 base pairs. Most of these are single nucleotide polymorphisms (SNPs) (Figure 1a) that can be detected in various ways. For example, many SNPs alter sites cleaved by restriction enzymes and can be used as cleaved-amplified polymorphic sequence (CAPS) (Konieczny and Ausubel, 1993). Alternatively, derived CAPS (dCAPS) is an elegant way to introduce a



**Figure 1.** Examples of single nucleotide polymorphism (SNP) and insertion-deletion (InDel) polymorphisms. Two markers from the Cereon *Arabidopsis* Polymorphism Collection are shown.

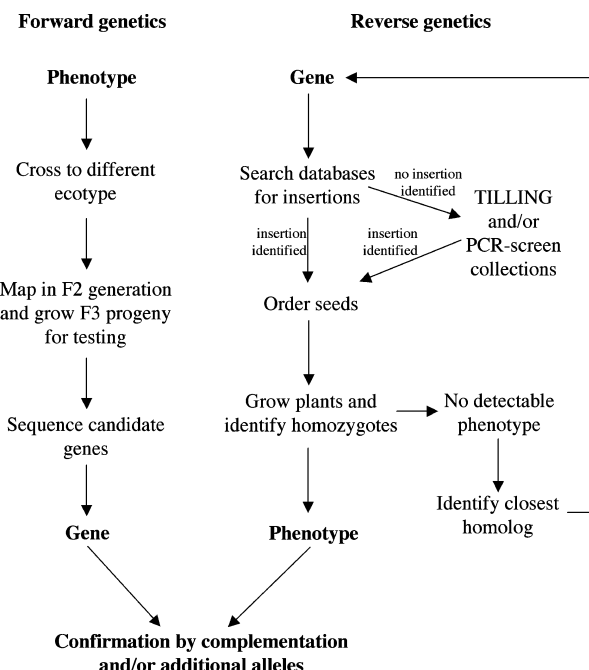
(a) Marker 447439 has a single-nucleotide change from T in Col-0 to C in Ler. This generates an *HpaII* restriction site in the Ler genome (underlined). (b) Marker 450823 has an eight-nucleotide insertion in Col-0 compared with Ler.

restriction recognition site in a primer containing the SNP, when the SNP itself does not provide a restriction site (Neff *et al.*, 1998). Also, a single-strand conformational polymorphism (SSCP) strategy can be employed to detect mismatches between wild type and mutant DNA by electrophoresis (Nataraj *et al.*, 1999). Access to the Cereon collection is obtained by one-time registration through The *Arabidopsis* Information Resource (TAIR) website (<http://www.arabidopsis.org/Cereon>). Besides SNPs, this collection also contains a list of insertion-deletion (InDel) markers (Figure 1b), which can be readily detected by PCR and appropriate gel electrophoresis techniques.

Given a sequenced genome and a dense collection of genetic markers, map-based cloning has become greatly facilitated compared with the previous chromosome walking process. With today's amount of markers on the *Arabidopsis* genome, it is estimated that the minimal start-to-finish time for a mapping project is one person-year (Jander *et al.*, 2002). This includes five generations of plant growth. A simple outline of the map-based cloning process with subsequent confirmation of linkage between gene and phenotype is shown in Figure 2 (left part and bottom). For further descriptions of map-based cloning protocols and projects, excellent reviews have recently been published on this subject (Jander *et al.*, 2002; Lukowitz *et al.*, 2000).

### Reverse genetics

Whereas forward genetics starts with the mutant and then leads to the gene, reverse genetics starts with the gene of interest and ends with the corresponding mutant. This



**Figure 2.** Overview of the process of establishing gene function by mutagenesis. To the left, a simplified outline for a forward genetics approach is shown (for further details of the map-based cloning process, see Jander *et al.* (2002) or Lukowitz *et al.* (2000)) and to the right the individual steps in a reverse genetics approach is shown.

approach is particularly useful for dissecting gene families where functional redundancy among closely related gene members often obscures their phenotypes. To circumvent this problem, reverse genetics allows mutations in all members of a gene family to be identified. Subsequently, simple crosses can be performed to combine mutations in closely related genes, often revealing the function of genes that would otherwise remain hidden. In many model systems, including prokaryotes, yeast, and even mice, gene targeting by homologous recombination is the method of choice for inactivating a gene of interest. However, with the exception of a few isolated examples of similar procedures in plants (Beetham *et al.*, 1999; Kempin *et al.*, 1997), no routine and efficient method for homologous gene targeting is currently available for plants. Instead, researchers have created near-saturation libraries of insertion mutant alleles that can be easily screened by polymerase chain reaction (PCR) for insertions into the gene of interest. More recently, the sequencing of the insertion sites for a large number of these insertion lines has enabled computer searches for mutations in the gene of interest. While it may take some time before such insertion libraries reach saturation, it is now possible to screen *in silico* for a mutation in the gene of interest, order the seeds from the *Arabidopsis* stock center, and begin analyses of mutant phenotypes within days. These remarkable advances in technology and resources are dramatically altering the way in which plant geneticists carry out their

research and should soon allow for the unmasking of mutant phenotypes among gene families. Another alternative to these approaches is to down-regulate the gene of interest through antisense or co-suppression (Brusslan *et al.*, 1993) or the more recent refinements of these methods using RNAi-based technology. RNAi allows for targeted down regulation of genes, and vectors such as pHANNIBAL have been developed in which inverted repeats of a gene sequence can be inserted (Wesley *et al.*, 2001). A genome-wide RNAi project based on transient expression was recently reported for *Caenorhabditis elegans* (Kamath *et al.*, 2003). Such an undertaking has not been initiated in *Arabidopsis*, most likely because of lack of transient expression systems and difficulties with inheritance of stable phenotypes.

Even with near-saturation insertion collections, success of obtaining an observable mutant phenotype by identifying a knockout of a particular gene is not guaranteed. The genome-scale RNAi approach in *C. elegans* by Kamath and Ahringer (2003) revealed that phenotypes could only be obtained in about 10% of the 16 757 genes studied. This relatively low number reflects the fact that most genes in higher organisms are members of gene families and can act redundantly. Meinke *et al.* (2003) estimate that of all the predicted *Arabidopsis* genes, about 65% appear to be members of families with two or more members, and given this level of redundancy in *Arabidopsis*, and based on phenotype detection in *C. elegans*, only about 10% of *Arabidopsis* genes (less than 3000) are expected to result in a detectable loss-of-function phenotype (Meinke *et al.*, 2003). It may therefore be necessary to construct double, triple or even higher-order knockouts to reveal the hidden phenotypes (Liljégren *et al.*, 2000; Pelaz *et al.*, 2000). In addition, some phenotypic characteristics may be hard to detect unless the mutated gene is studied in a certain mutant background that more clearly reveals its loss-of-function phenotype (Roeder *et al.*, 2003). In other cases, mutations in one gene, which has no detectable mutant phenotype by itself, can lead to suppression of a phenotype from a mutation in another gene (Li *et al.*, 1999). Alternatively, mutagenizing a mutant line of interest can uncover the function of an unrelated gene with an overlapping function (Eshed *et al.*, 1999). Assessing a possible phenotype may also depend on the assay conditions. For instance, mutations in genes that are involved in stress responses may only display a detectable phenotype when subjected to certain environmental challenges. Reverse genetics largely facilitates these kinds of studies by allowing scientists to make qualified guesses on which combinations of mutations would give rise to phenotypic changes.

The following sections will cover the process of obtaining mutants through reverse genetics, and review recent progress in the application of large-scale insertional mutagenesis for functional genomics in *Arabidopsis*.

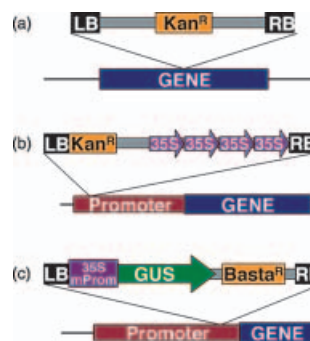
## T-DNA tagging

*Agrobacterium tumefaciens* has often been called 'nature's genetic engineer' because of its natural ability to transfer a segment of its DNA into plant genomes. This T-DNA segment is flanked by short imperfect direct repeat border sequences called left and right T-DNA borders (reviewed in Zambryski, 1988). As indicated in Figure 3(a), the T-DNA is defined as any sequence between these two borders and can be directed to integrate into plant genomes in a largely random manner.

The *Agrobacterium*-mediated floral dip method has become the most widely used method of generating transgenic *Arabidopsis* plants (Bechtold and Pelletier, 1998; Bechtold *et al.*, 1993; Clough and Bent, 1998). The ability to transform plants with foreign DNA has revolutionized plant research and agriculture. It is now possible to have plants expressing essentially any gene of interest for basic research or industry applications. Moreover, the largely random nature of T-DNA insertion distribution throughout the *Arabidopsis* genome has made it a powerful system for large-scale insertional mutagenesis (Alonso *et al.*, 2003; Azpiroz-Leehan and Feldmann, 1997; Sessions *et al.*, 2002; Sussman *et al.*, 2000) as well as for activation tagging and enhancer trap studies (see below). A major advantage of using T-DNA as an insertional mutagen is that it provides a molecular tag that greatly facilitates isolation of the corresponding gene.

### Finding a T-DNA insertion in the gene of interest

The process of finding a mutant plant with a knockout of a gene of interest has become more straightforward with the advent of near-saturation T-DNA collections. Data on the



**Figure 3.** Transferred-DNA (T-DNA) insertions.

(a) Schematic illustration of insertion of a simple T-DNA construct with left (LB) and right (RB) borders and carrying a Kanamycin resistance gene (Kan<sup>R</sup>) into a random gene on a chromosome.

(b) An example of an activation-tagging construct carrying a cassette with four copies of the 35S enhancer from cauliflower mosaic virus (Benfey and Chua, 1989) pointing in the direction of the RB. Insertion into the promoter region is shown.

precise chromosomal location of individual inserts have been obtained largely by thermal asymmetric interlaced (TAIL) PCR technology (Liu and Whittier, 1995), subsequent sequencing of the resulting fragments, and deposition in databases. Users can now do *in silico* searches in these databases for lines in which a T-DNA has been inserted in a particular location of the genome. Sessions *et al.* (2002) were the first to report the generation and thorough characterization of a relatively large insertion collection containing 52 964 T-DNA lines with flanking sequence and 15 000–18 000 of the *Arabidopsis* genes having insertions (The SAIL collection; Table 1). At the Salk Institute, more than 225 000 independent insertion events have been created, and the precise location has been determined for T-DNA insertions in approximately 90 000 lines (Alonso *et al.*, 2003) (The SIGnAL collection; Table 1). These and other T-DNA insertion databases are available to the public and can be searched on the websites shown in Table 1. In this section, we will discuss (i) how to identify the lines, (ii) how to order them from the stock center, and (iii) how to initially grow and genotype the plants to identify which are homozygous for the mutation. See also Figure 2 for a schematic outline of the process of reverse genetics in comparison with forward genetics discussed above.

1. When one is interested in studying the loss-of-function of a particular *Arabidopsis* gene and has decided to address this by taking a reverse genetics approach, the first step is to enter an insertion collection website to search for an insertion in the gene of interest. We suggest, initially, to perform searches at the *Arabidopsis thaliana* Insertion Database (AtIDb) (Pan *et al.*, 2003) or through the Nottingham *Arabidopsis* Stock Centre (NASC) websites (Table 1). At these websites insertion data from a variety of collections including several of the ones presented in Table 1 are compiled. One simply enters the gene accession number in the provided field and submits. These numbers are now standardized in the AtNGxxxxx format, which are called the AGI codes and are assigned by the *Arabidopsis* Genome Initiative to genes, either known or computationally predicted. At refers to *Arabidopsis thaliana*, N refers to the chromosome number, and xxxxx is a number given based on the position of a gene location on that particular chromosome. If mutations are found, a list of insertion lines is provided stating which collections they originate from.

Alternatively, with a direct search in the SIGnAL collection (Table 1), a map is provided which indicates the position of the gene, location of the nearest insertions, and orientation of these insertions as well. Moreover, it is also possible to access the flanking sequence that was obtained. This can give valuable information on how well the tagged sequence aligns with the sequence around the proposed insertion site.

2. If you do find a 'hit' in any of the collections, the next step is to order a batch of seeds derived from that insertion line from the *Arabidopsis* Biological Resource Center (ABRC)

at Ohio State University or the Nottingham *Arabidopsis* Stock Centre (NASC) using their websites or from the insertion collection itself (Table 1). Also seeds from the SAIL lines can be ordered from ABRC from May 2004.

3. The stock center sends out a small number of seeds for each insertion line (~100). These seeds are often segregating for the insertion (25% homozygous wild type:50% hemizygotes:25% homozygous mutant), so that the users will have to isolate homozygous mutant plants themselves. Kanamycin or glufosinate (Basta; AgrEvo, Frankfurt, Germany) resistance is encoded on most T-DNA vectors, however, it often happens that these resistance genes are suppressed in the transgenic lines. We therefore recommend not using these selections when planting the seeds. We find that the most efficient way of genotyping is to sterilize and plate the seeds, transfer seedlings to soil, and isolate genomic DNA as described in the General methods. Two PCR reactions are needed for each preparation (Figure 4): One reaction includes two gene-specific oligonucleotide primers (GS1 and GS2 in Figure 4a) to detect the wild-type allele, and another reaction with one gene-specific primer and one T-DNA primer (LB in Figure 4a) to detect the mutant allele. As genotyping using right border primers has occasionally been problematic because of several examples of T-DNA insertions with left borders on both ends (see FAQs section on SIGnAL website), we recommend using left border-specific primers when possible. T-DNA left border primer sequences that are routinely used are listed in Table 2. Which gene-specific primer to include with the T-DNA primer depends on the orientation of the T-DNA insert in the genome. This orientation is usually given on the website where it was originally identified. It is a good idea to include reactions of wild-type DNA as well for comparison. Figure 4(b) provides an example of a T-DNA genotyping experiment with nine plants segregating for insertion in the *HARMLESS TO OZONE LAYER (HOL)* gene (Rhew *et al.*, 2003). Plants were tested for the presence of the wild type allele (upper gel) and mutant allele (lower gel). In this particular example plant nos. 1, 8, and 9 are homozygous for the *hol-1* mutation (SIGnAL line #SALK\_005204).

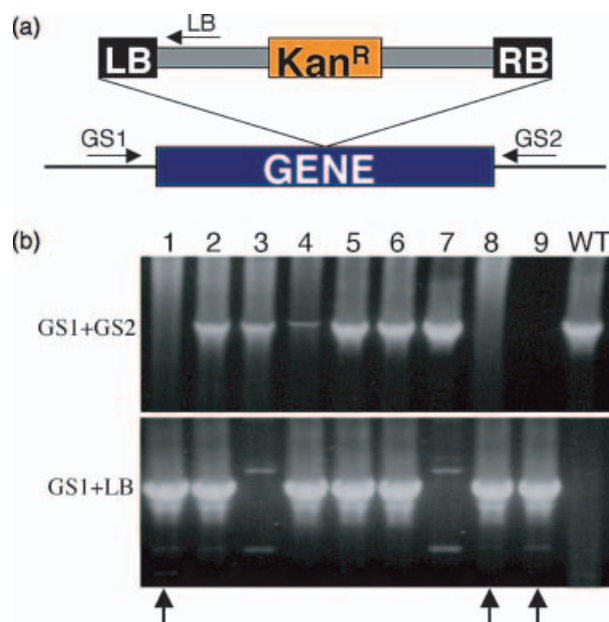
It should be noted that if one is working on a gene for which a homozygous mutation is lethal, one will not achieve a simple 1:2:1 band pattern of homozygous wild type:heterozygous:homozygous mutant, respectively. Instead, a 1:2 segregation of homozygous wild type:heterozygous is expected.

After having identified a plant with an insertion in the gene of interest, we recommend determining the number of insertions in the plant by Southern blotting with digested genomic DNA from the mutant using a fragment of the T-DNA as a probe. By using restriction enzymes that cut outside the T-DNA, only one band should be observed in each lane. If more than one insertion is detected, it is necessary to backcross the line to wild type until only the

**Table 1** *Arabidopsis* initiatives, available insertion collections, and stock centers

General <i>Arabidopsis</i> initiatives	Web addresses	No. of lines	Ecotype	Reference
The <i>Arabidopsis</i> Information Resource (TAIR)	<a href="http://www.arabidopsis.org">http://www.arabidopsis.org</a>			
The MIPS plant genomics group	<a href="http://www.mips.gsf.de/proj/thal/">http://www.mips.gsf.de/proj/thal/</a>			
GARNet	<a href="http://www.garnet.arabidopsis.org.uk">http://www.garnet.arabidopsis.org.uk</a>			
Gene annotations and insertion information from a variety of insertion collections				
<i>Arabidopsis thaliana</i> insertion database (AtIDb)	<a href="http://www.atidb.cshl.org">http://www.atidb.cshl.org</a>			Pan <i>et al.</i> (2003)
NASC	<a href="http://www.arabidopsis.org.uk/blast.html">http://www.arabidopsis.org.uk/blast.html</a>			
Gene/ORFeome Information				
SSP consortium	<a href="http://www.signal.salk.edu/SSP/index.html">http://www.signal.salk.edu/SSP/index.html</a>			Yamada <i>et al.</i> (2003)
The TIGR <i>Arabidopsis thaliana</i> Genome Annotation Database	<a href="http://www.tigr.org/tdb/e2k1/ath1/ath1.shtml">http://www.tigr.org/tdb/e2k1/ath1/ath1.shtml</a>			
T-DNA Insertion Collections with identified T-DNA insertion				
Syngenta <i>Arabidopsis</i> Insertion Library (SAIL)				
The SIGNAL collection	<a href="http://www.tmri.org">http://www.tmri.org</a>	52 964 <sup>a</sup>	Col-0	Sessions <i>et al.</i> (2002)
GABI-Kat	<a href="http://www.signal.salk.edu/cgi-bin/tdnaexpress">http://www.signal.salk.edu/cgi-bin/tdnaexpress</a>	90 000 <sup>a</sup>	Col-0	Alonso <i>et al.</i> (2003)
The SeedGenes database	<a href="http://www.mpiz-koeln.mpg.de/GABI-Kat/">http://www.mpiz-koeln.mpg.de/GABI-Kat/</a>	44 358 <sup>a</sup>	Col-0	
Versailles FST project	<a href="http://www.seedgenes.org">http://www.seedgenes.org</a>	218 <sup>a</sup>	Col-0	Tzafrir <i>et al.</i> (2003)
	<a href="http://www.flagdb-genoplante-info.infobiogen.fr/">http://www.flagdb-genoplante-info.infobiogen.fr/</a>	28 998 <sup>a</sup>	WS	Balzerque <i>et al.</i> (2001)
Transposon Insertion Collections				
Gene/Enhancer trap collection at CSHL	<a href="http://www.genetrap.cshl.org/">http://www.genetrap.cshl.org/</a>	16 887 <sup>a</sup>	Ler	Martienssen (1998)
Sainsbury Laboratory <i>Arabidopsis</i> Transposants (SLAT)	<a href="http://www.arabidopsis.info/info/slat_info1.html">http://www.arabidopsis.info/info/slat_info1.html</a>	1000 <sup>a</sup>	Col-0	
Collections for gene-specific screening				
The <i>Arabidopsis</i> knockout facility (Alpha)	<a href="http://www.biotech.wisc.edu/Arabidopsis">http://www.biotech.wisc.edu/Arabidopsis</a>	60 480	Ws	Sussman <i>et al.</i> (2000)
The <i>Arabidopsis</i> knockout facility (Basta)	<a href="http://www.biotech.wisc.edu/Arabidopsis">http://www.biotech.wisc.edu/Arabidopsis</a>	72 960	Ws	Sussman <i>et al.</i> (2000)
The SIGNAL collection	<a href="http://www.signal.salk.edu/cgi-bin/tdnaexpress">http://www.signal.salk.edu/cgi-bin/tdnaexpress</a>	225 000	Col-0	Alonso <i>et al.</i> (2003)
SLAT	<a href="http://www.arabidopsis.info/info/slat_info1.html">http://www.arabidopsis.info/info/slat_info1.html</a>	40 000	Col-0	
TILLING <sup>b</sup>	<a href="http://www.proweb.org/input/">http://www.proweb.org/input/</a>		Col-0	Till <i>et al.</i> (2003)
Enhancer/promoter trap and activation tagging collections				
GAL4-GFP enhancer trap lines	<a href="http://www.plantsci.cam.ac.uk/Haseloff/">http://www.plantsci.cam.ac.uk/Haseloff/</a>	>8000	C24	Kiegle <i>et al.</i> (2000)
GAL4-GFP enhancer trap lines	<a href="http://www.enhancertraps.bio.upenn.edu/">http://www.enhancertraps.bio.upenn.edu/</a>		Col-0	
The RIKEN Activation Tagging Line Database	<a href="http://rage.gsc.riken.go.jp/activationtag/top.php">http://rage.gsc.riken.go.jp/activationtag/top.php</a>	1172	Col-0	Ichikawa <i>et al.</i> (2003)
Enhancer trap collection at CSHL	<a href="http://www.genetrap.cshl.org/">http://www.genetrap.cshl.org/</a>	14 204	Ler	Martienssen (1998)
Stock Centers				
<i>Arabidopsis</i> Biological Resource Center (ABRC)	<a href="http://www.arabidopsis.org/abrc/">http://www.arabidopsis.org/abrc/</a>			
Nottingham <i>Arabidopsis</i> Stock Centre (NASC)	<a href="http://www.nasc.nott.ac.uk/">http://www.nasc.nott.ac.uk/</a>			

<sup>a</sup>Number of lines with flanking sequence.<sup>b</sup>*er105* fast neutron-induced mutant three times backcrossed to Col-0 (Torii *et al.*, 1996).



**Figure 4.** Example of genotyping of segregating plants of a Transferred-DNA (T-DNA) insertion line.

(a) Schematic drawing of the insertion with position of primers. #GS1 is gene-specific primer 1, #GS2 is gene-specific primer 2, and #LB symbolizes a primer in the T-DNA close to the left border.

(b) Agarose gel electrophoresis to visualize PCR result of genotyping the *hol-1* allele (Rhew *et al.*, 2003). Upper panel represents PCR reactions including #GS1 and #GS2 to monitor the presence of a wild type allele. Lower panel are PCR reactions with #GS1 and #LB for detection of mutant allele. Arrows point to reactions (#s 1, 8, and 9) that show the presence of homozygous *hol-1* mutant plants. WT represents reactions with wild type genomic DNA as control.

desired insertion is present in order to rule out effects of secondary mutations. After verifying that the insertion is indeed within the gene of interest it is then necessary to determine if the insertion effects the transcript levels and/or size. To achieve this result, it is generally recommended that RNA from homozygous mutant plants be analyzed by

Northern blotting using a gene specific probe that is 3' of the insertion site. Numerous reports describing T-DNA mutant characterizations have been published within the last few years, and a recent example was provided by Rhew *et al.* (2003).

#### When database searches fail to identify insertions

Even with large collections of several hundred thousand lines, it is possible that an insertion in the gene of interest may not have been reported. The probability,  $P$ , of finding an insertion in a given gene in a given population assuming random distribution can be calculated according to the formula

$$P = 1 - [1 - (X/125\,000)]^n$$

where  $X$  is the length of gene in kilo base pairs, 125 000 corresponds to the approximate size of the *Arabidopsis* genome in kilo base pairs, and  $n$  is the number of T-DNA inserts present in the population (Krysan *et al.*, 1999). Therefore, if one is working on a small gene of <1000 bp the probability of an insertion within that gene is relatively low (<~60% for collections with 100 000 lines). Because only a fraction of the available T-DNA and transposon insertion collections have had their insertion sites sequenced, there are several large collections that are publicly available for researchers to conduct their own PCR-based screens. For example, it is possible to obtain genomic DNA preparations from the SiGnAL collection through the ABRC stock center.

The first step in such screens involves PCR amplification using a gene-specific primer and a primer specific for the insertion element (e.g. T-DNA or transposon) using a set of DNA preparations containing pools of genomic DNA from different insertion lines. Pools are constructed to minimize the number of PCR reactions necessary to identify the seeds harboring the desired insertion (for descriptions of pool architecture, see e.g. Krysan *et al.*, 1999; Sussman *et al.*,

**Table 2** T-DNA and transposon specific primers

T-DNA <sup>a</sup>	Position <sup>b</sup>	Sequence	Collection
JMLB1	90-61	5'-GGC AAT CAG CTG TTG CCC GTC TCA CTG GTG-3'	SiGnAL
JMLB2	341-312	5'-TTG GGT GAT GGT TCA CGT AGT GGG CCA TCG-3'	SiGnAL
VLB1	166-143	5'-CGG CTA TTG GTA ATA GGA CAC TGG-3'	Versailles
LB1	~300	5'-GCC TTT TCA GAA ATG GAT AAA TAG CCT TGC TTC C-3'	SAIL
JL202	166-138	5'-CAT TTT ATA ATA ACG CTG CGG ACA TCT AC-3'	<i>Arabidopsis</i> Knockout Facility
JL270	70-42	5'-TTT CTC CAT ATT GAC CAT CAT ACT CAT TG-3'	<i>Arabidopsis</i> Knockout Facility
Ds primers	Bp from border	Sequence	Collection
DS3.3	123-101	5'-GTA TTT ATC CCG TTC GTT TTC GT-3'	Cold Spring Harbor Gene trap
DS3.4	76-56	5'-CCG TCC CGC AAG TTA AAT ATG-3'	Cold Spring Harbor Gene trap

T-DNA and transposon specific primers that have been successfully used in the Yanofsky laboratory for genotyping.

<sup>a</sup>Only primers proximal to the left border sequences are listed.

<sup>b</sup>Distance in base pairs from border between insertion and tag.



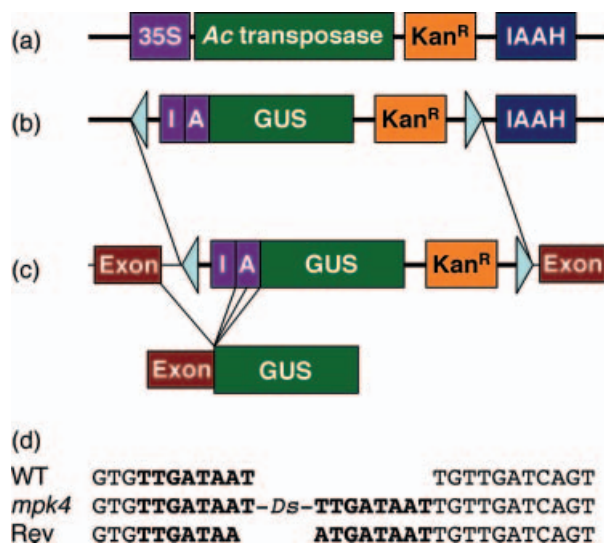
2000; Winkler *et al.*, 1998). The second step involves Southern blotting these PCR products using a probe specific for the gene of interest. Positive signals in a certain set of reactions will then allow for the identification of the line, and seeds from that line can subsequently be ordered from the stock centers.

The *Arabidopsis* knockout facility at the University of Wisconsin has established a user-fee service facility at <http://www.biotech.wisc.edu/Arabidopsis> to provide knockout *Arabidopsis* mutant plants for the plant research community (Sussman *et al.*, 2000). The users must provide primers that have been tested in control amplifications. The facility will then do the PCR and send the reactions back to the customer for Southern blotting and hybridization analysis. The *Arabidopsis* knockout facility has two collections. One is called the Alpha collection and contains 60 480 independent lines, and the other is called the BASTA collection and has 72 960 lines (Table 1). The T-DNA vector used in the Alpha collection carries a small portion of the *APETALA3* promoter, which occasionally leads to co-suppression of the *AP3* gene and subsequent fertilization problems (Krysan *et al.*, 1999). We recommend visiting the facility website for further description of the collections and how to set up a screen.

To save time and money, we suggest that the SIGnAL collection and the *Arabidopsis* knockout facility collections should be screened manually only after having checked the databases for insertions in your gene of interest. It is worth noting that before discarding the possibility of an insertion in a particular gene, one should carefully check the annotation of the gene done by the various genome databases, as these efforts can have a fairly high degree of misannotation.

### Transposon tagging

Transposable elements mutagenize genes by insertion into coding and regulatory regions, and researchers have utilized this system for functional genomics initiatives in *Arabidopsis* and other plants (Martienssen, 1998; Parinov and Sundaresan, 2000; Speulman *et al.*, 1999; Sundaresan *et al.*, 1995). Sundaresan *et al.* (1995) reported the use of the maize-derived *Activator/Dissociation* (*Ac/Ds*) system in *Arabidopsis* to create a gene disruption collection (gene trap) and a collection to identify *cis*-acting regulatory elements (enhancer trap). These public collections include, at the time of writing, 19 237 gene trap lines and 14 262 enhancer trap lines. In total, 16 887 lines with sequenced flanking regions are currently available (The CSHL collection in Table 1). The insertion lines were created by crossing two separate starter lines. One carrying a T-DNA with a DNA fragment encoding the *Ac* transposase and one harboring a T-DNA insertion with the non-autonomous *Ds* element (Figure 5). The two lines were crossed to allow transposition events to take place, and the *Ac* transposase was subsequently crossed out



**Figure 5.** Schematic overview of the *Ac/Ds* transposon system.

(a) *Ac* element encoding transposase in T-DNA vector. (b) The gene trap element, *DsG*, in T-DNA vector.

(c) Expression from *DsG* inserted into an intron.

(d) Example of transposon insertion and reversion in the *mpk4* gene (Petersen *et al.*, 2000). Notice that eight base pairs 5' of the insertion site are repeated at the 3' of the insertion. Notice also that the transposon leaves a 'footprint' when excised. IAAH: indole acetic acid hydrolase gene conferring sensitivity to NAM; I: *Arabidopsis* intron sequence; A: Triple acceptor splice site. Light blue arrowheads indicated the borders of the *Ds* element. GUS is the  $\beta$ -glucuronidase reporter gene, and Kan<sup>R</sup> encodes the NPTII gene providing kanamycin resistance.

to prevent further mobilizations (Sundaresan *et al.*, 1995). For the CSHL collection, the *Ds* element was cloned in the T-DNA vector in combination with either a minimal promoter in front of the  $\beta$ -glucuronidase (GUS) reporter gene (for enhancer trapping) or an *Arabidopsis* intron with acceptor site in all three reading frames followed by the GUS gene as well (for gene trapping). An overview of the gene trap system is shown in Figure 5(a–c).

Whereas T-DNA produces stable immobile integrations, transposons can often be remobilized from the insertion site to revert a potential phenotype. Such remobilization can provide important confirmation about mutational effects of insertions (Petersen *et al.*, 2000) (Figure 5d). Analyzing >500 lines revealed two 'hot spots' for *Ds* insertions – one on chromosome 2 and one on chromosome 4. Apart from these, distribution appeared evenly throughout the genome (Parinov *et al.*, 1999).

The unstable nature of transposon integration can, however, also be a caveat for using this system in large-scale mutagenesis. Cases have been reported for which a phenotype was not linked to the *Ds* element. One way this may happen is if the *Ds* element jumped more than once in the *Ac/Ds* F1 plants, leaving footprints inside coding sequences and thereby altering the reading frame and changing the gene product. It has also been reported that transposition

can create large chromosomal deletions upon mobilization. In one such example (Brodersen *et al.*, 2002), the transposable unit excised a 30 kb fragment giving rise to a deletion of five genes.

A facility at Sainsbury Laboratory at the John Innes Centre, Norwich, UK, has developed nylon filter-bound arrays with the representation of more than 40 000 transposon insertions in *Arabidopsis* (May *et al.*, 2002) (Table 1). These so called SLAT filters (Sainsbury Laboratory *Arabidopsis* Transposants) are presented in an array of 864 spots each containing PCR-amplified products of flanking sequence for 50 transposant lines. Following hybridization with suitable probes a single pool of 50 lines may be identified which can then be further screened.

The number of transposon lines in the public databases is not as high as it is for T-DNA lines. Nevertheless, web addresses for transposon insertion collections are provided in Table 1, and in case of positive 'hits' in your gene of interest, seeds can be obtained from the stock centers as well. Isolation of homozygous lines can be done as described for T-DNA lines, and we have had good success using the primers DS3.3 and DS3.4 (Table 2) in PCR reactions for genotyping *Ds* insertions.

As for T-DNA insertion mutants, it is important to establish whether the phenotype is in fact the result of the integration of the transposable element into the gene of interest. Considerations regarding the relevance of the phenotype may give an idea (is the gene even expressed in the affected tissue of a wild-type plant?), but in addition, we strongly recommend Southern blot and Northern blot analyses as described above for T-DNA insertions. These experiments will show if there are more insertions, and if expression of the gene is affected. For further evidence regarding confirmation of any kind of mutation, see below under 'Confirmation of mutant phenotype.'

## TILLING

The near-saturation insertion collections in *Arabidopsis* have proven useful in a number of cases to markedly improve our understanding of a subset of *Arabidopsis* genes. However, as projects approach completion targeted methods will be necessary to reach the goal of knocking out all the genes. This is especially the case for small genes where huge numbers of lines are needed to obtain a high probability of achieving an insertion mutant (see above).

Targeting Induced Local Lesion in Genomes (TILLING) is a general reverse-genetics tool, which combines random chemical mutagenesis with PCR-based screening to identify point mutations in a genomic region of interest (McCallum *et al.*, 2000). Seeds from one plant were mutagenized by EMS, and a population of 6912 arrayed DNAs from mutagenized individuals is presently available for screening (Till

*et al.*, 2003). Note that this number is significantly higher than the 5000 M1 plants assumed to be enough for saturated EMS mutagenesis screens (Feldmann *et al.*, 1994). The more M1 plants, the more likely it is to obtain several changes within a single gene, which may lead to more complete allelic analyses.

Regions of interest spanning 1000 base pairs are defined by the user and are amplified by PCR at the *Arabidopsis* TILLING Project (ATP) facility. Mismatched base pairs in heteroduplex formations between wild-type and mutant fragments are recognized and cleaved by the CEL1 nuclease. Subsequent denaturing polyacrylamide gel electrophoresis and sequencing of the fragments allows for identification of mutant seeds stocks that can be obtained from the stock centers (Colbert *et al.*, 2001).

Plant researchers can obtain mutants using the TILLING technology in five steps, and start the procedure by going to <http://www.proweb.org/input> to obtain gene models and protein conservation models. An easily comprehensible step-wise description of the subsequent process is provided by Till *et al.* (2003). The ATP facility has a policy of immediate public release of data, and mutations in 1-kb regions already screened can be identified at <http://tilling.fhcrc.org:9366/search.html>. Therefore, before ordering a TILLING project of your gene of interest, it is worthwhile checking if mutations are already available in that region.

## Confirmation of mutant phenotype

After having isolated a T-DNA, transposon, or TILLING line with an insertion or point mutation in the gene of interest, the next step is to determine if the mutation results in a phenotype that can be distinguished from wild type, thereby pointing to a function of that gene. However, even if this is the case, it is necessary to confirm that the gene is indeed responsible for that phenotype (Figure 2). T-DNA lines of several collections contain an average of 1.5 insertions per line (Alonso *et al.*, 2003; Feldmann, 1991), and it is therefore possible that an insertion somewhere else in the genome is responsible for the phenotype. Furthermore, second site mutations can arise that are not T-DNA tagged. In addition, the unstable nature of transposons described above makes it important to establish that the correct gene is being investigated.

Perhaps the fastest method for demonstrating that the mutant phenotype is the result of a mutation in the gene of interest is to identify additional mutant alleles that have been independently isolated. As noted above, the near-saturation collections of insertion mutants in *Arabidopsis* makes this a straightforward process, where one searches *in silico* for new insertion alleles, orders them from the stock center, and then analyzes the resulting plants for the same mutant phenotypes. Having several independently isolated mutant alleles that produce similar phenotypes is generally

accepted as convincing evidence that the mutated gene is indeed responsible for the mutant phenotype. The more labor-intensive method, however, which even today is held as the standard procedure to prove that the mutated gene is responsible for the phenotype, is to perform a gene rescue experiment. For this approach, the wild-type copy of the gene, including all introns, exons, about 3 kb of 5' flanking promoter sequence, and around 1 kb 3' flanking sequence is transformed directly into the mutant. In cases where the homozygous mutants are lethal or unable to produce progeny, transformations are carried out on heterozygous individuals. If the wild-type transgene is able to fully rescue the mutant phenotype, it is safe to conclude that the correct gene has been identified. Some researchers have used a somewhat less satisfying approach. In these studies, they have fused the cDNA for the gene of interest to the strong constitutive cauliflower mosaic virus (CaMV) 35S promoter (Benfey and Chua, 1989) and have asked whether or not this transgene is able to rescue the mutant phenotype. While such an experiment can be informative, it can also lead to misinterpretations. For example, overexpression of a closely related gene may be able to rescue the mutant phenotype even if the mis-expressed gene is not the one that is mutated.

### Effects of mutations

It is not always straightforward to evaluate the strength of a specific mutation in a gene of interest. However, for T-DNA and transposon insertions it is generally considered that the further upstream in the gene the insertion occurred the more likely it is to cause complete loss-of-function (knockout). Moreover, insertions within exons are preferable because the T-DNA or transposon DNA might be spliced out when inserted in an intron. As stated above, expression levels can be checked by Northern blot analysis using a probe which is specific to a sequence 3' of the insertion site.

For chemical or fast neutron mutagenesis, introduction of stop codons or small insertions or deletions that disrupt the reading frame will typically be more severe the further upstream in a gene they occur. It therefore follows that the smaller the resulting truncated protein product, the less likely it is to retain any activity. However, such alterations frequently do not affect expression levels and Northern blot analysis is therefore not an efficient method to evaluate the strength of, for example, an EMS allele. Instead, it is necessary to monitor the phenotype either of the single mutant or in combination with mutations in genes with redundant functions.

### Assisting functional genomics approaches

Activation tagging is an alternative approach to the conventional loss-of-function mutagenesis described above for

functional genomics studies. Whereas chemical mutagens, transposons, and T-DNA insertions mainly create recessive mutations, activation tagging can produce dominant or semi-dominant phenotypes so that the effect of wild-type loci can be largely ignored (Marsch-Martinez *et al.*, 2002; Weigel *et al.*, 2000). About a decade ago Hayashi *et al.* (1992) constructed a T-DNA vector with four copies of the enhancer elements of the CaMV35S promoter (Figure 3b). These enhancers can transcriptionally activate nearby genes and provide a tag for the identification of the chromosomal location by, for example, TAIL-PCR or plasmid rescue. If the activation construct integrates within a gene, it may create a knockout. However, it can also result in increased expression of the region downstream of the insertion site leading to phenotypes that may be different from complete loss-of-function of that particular gene. Several groups have achieved interesting phenotypes that have led to important information regarding the function of a number of genes (Dinnyen *et al.*, 2004; Kardailsky *et al.*, 1999; Nakazawa *et al.*, 2003; Palatnik *et al.*, 2003). Ichikawa *et al.* (2003) reported the screening of 55 431 activation-tagged lines of which 1262 showed phenotypes different from wild type. The chromosomal locations were established by plasmid rescue for 1172 of those, and they can be searched by the plant science community at <http://range.gsc.riken.go.jp/activationtag/top.php> (Table 1). While the Northern blot analysis is frequently successful at identifying the gene of interest, Ichikawa *et al.* (2003) describe an instance in which 8 kb separate the enhancer from the target gene and examples of fruitless searches are not uncommon.

Other functional genomics methodologies have been developed to increase our insight into gene function. Knowledge about expression pattern often provides a first hint that a particular gene may have a function in a particular process. The study of differential expression patterns of a huge number of genes in response to a number of stimuli has become possible with development of the micro-array technique. A full genome *Arabidopsis* chip is available from Affymetrix, and enormous amounts of data can be created in a relatively short time, corresponding to tens of thousands of Northern blots in every experiment. The micro-array technology is still costly, requiring expensive chips and expensive equipment, but facilities that will perform the chip hybridization and reading have become abundant, so that the user has 'only' to provide the chip and RNA.

For more tissue-specific expression information, enhancer trap lines are useful and an example of a general construct is shown in Figure 3(c). A transposon-based approach was initiated at Cold Spring Harbor a few years ago (Martienssen, 1998) (Table 1), and Haseloff and colleagues have built an enhancer/promoter trap collection based on the GAL4-GFP reporter system, which can be screened on their website (Kiegle *et al.*, 2000) (Table 1).

These collections provide the user with tissue-specific expression, and seed from individual lines can be obtained through the stock centers (Table 1).

### Perspectives

In the past 50 years, biological research has been revolutionized because of several astonishing discoveries including, for example, the structure of DNA (Watson and Crick, 1953) and development of the PCR technology (Saiki *et al.*, 1988). Around the millennium, whole genomes of various organisms including *Arabidopsis*, *Caenorhabditis*, *Drosophila*, human, mouse, and several microorganisms were completely sequenced. These efforts have provided us with a wealth of data that has proven extremely useful for biological research in both functional and comparative genomics studies.

The analysis of mutants with detectable phenotypes is undoubtedly the source that has given us most insight into the mechanisms underlying a wide range of biological processes in plants. With the formation of genome-scale mutagenesis projects to assist the plant science community, it will most certainly continue to do so for the years to come. Obtaining mutations in genes of interest is an important and necessary step, and it has been our attempt to provide an overview of the different techniques that are currently at your disposal for the *Arabidopsis thaliana* model system. However, mutagenesis is just one of many powerful tools that are needed to understand how the function of a gene is carried out. Genome-wide micro-array analysis can be used as a powerful tool to provide information on which genes are involved in the formation of different cell types. For example, by comparing expression profiles of wild-type plants to mutants lacking the activity of a developmentally regulated transcription factor gene, it is possible to identify downstream genes of that transcription factor (Wellmer *et al.*, 2004). Moreover, the order and timing of gene activation, identification of protein interaction partners as well as biochemical characteristics of the enzymes carrying out the metabolic processes will prove crucial for us to fully understand the processes of *Arabidopsis* biology.

With the entire sequence of a reference plant genome available and a thorough understanding of the gene functions, we will be able to profoundly broaden our understanding of how other plants develop and behave by comparing the genes and their expression patterns. Such comparative genomics studies are indeed among the goals of the '2010 project' (Chory *et al.*, 2000). A first draft sequence of the rice genome sequence was recently published (Goff *et al.*, 2002), and it appears to be almost four times bigger than the *Arabidopsis* genome, but likely encoding fewer than twice the number of genes (~45 000 in rice compared with ~26 000 in *Arabidopsis*). T-DNA

insertion collections are presently being produced for rice (Chen *et al.*, 2003; Sha *et al.*, 2003), and comparing the functions of closely related genes allows us to ask questions within the exciting and rapidly expanding field of evolutionary and developmental biology (Evo-Devo). A major goal of Evo-Devo is to understand how developmental processes become modified during evolution and from these genetic changes to reveal how the past and present biodiversity arose. It will be interesting to study how members of gene families diverged to undertake different functions in different species (paralogy) or how a specific gene has maintained a similar function in otherwise distantly related species (orthology). Such cross-species analyses might therefore reveal those genes that are necessary for the development and maintenance of cells that are unique to diverse plant species and may enable us to eventually fulfill the goals of 'The 2010 project'.

### General methods

#### *Arabidopsis* growth conditions

Seeds on MS plates (Murashige and Skoog, 1962) or in soil are imbibed at 4°C for at least 2 days prior to transfer to a growth room. Our growth room conditions are 22°C with continuous light of 120–150  $\mu\text{mol m}^{-2} \text{sec}^{-1}$  of intensity.

#### Seed sterilization for plating on 80 mm Petri dishes with MS medium

Add a desired number of seeds (100–200) to of 15-ml Falcon tube. Wash with 70% ethanol for 2 min.

Remove ethanol.

Add bleach solution (2% NaClO, 0.01% Tween-20) and incubate 5–10 min.

Remove bleach solution and wash twice with sterile water.

Add 8 ml melted MS top agar and spread on MS plate.

For more details on *Arabidopsis* growth see Weigel and Glazebrook (2002).

#### Primer design

One of the most important factors for achieving success in a knockout screen is the selection of the PCR primers. It is very important to check that the primers conform to the guidelines provided by the facility that made the collection. For example, the *Arabidopsis* knockout facility at the University of Wisconsin (<http://www.biotech.wisc.edu/Arabidopsis>) describes the following guidelines for designing primers:

Length: 29 nucleotides, GC-content: 34–50%, and zero or one G or C at positions 28 and 29 (at the 3' end of primer).

#### Preparing genomic DNA for genotyping

1. Sterilize 10–20 seeds according to the protocol given above.
2. Place seeds on MS plates (without sucrose in the medium) (Murashige and Skoog, 1962)
3. Place the plates at 4°C for 4 days.

4. Transfer plates to growth room and incubate under standard plant growth conditions (see *Arabidopsis* growth conditions) for 5–7 days.
5. Transfer seedlings to soil and continue to grow under standard plant growth conditions.

When the seedlings have developed two pairs of leaves (not counting the cotyledons), one leaf can be taken and genomic DNA preparations can be prepared according to standard procedures (Edwards *et al.*, 1991).

We resuspend the DNA in 25 µl 10 mM Tris-HCl (pH 8.0) and use 5 µl in a 25 µl PCR reaction.

### Standard PCR conditions for genotyping

PCR conditions are as follows: denaturation at 95°C for 3 min, then 40 cycles of denaturation at 95°C for 15 sec; annealing at 55–60°C for 30 sec, and elongation at 72°C for 1–5 min. Annealing temperature will depend on the primer sequences that are used. The elongation time depends on the length of amplified fragment, but we usually allow 1 min per 1 kb.

### Acknowledgements

We would like to thank Chrystelle Asseman, Jose Dinnyen, Gary Ditta, Kristina Gremski, and Adrienne Roeder for critically reading the manuscript and providing fruitful comments. Our research is supported in part by a grant from the National Science Foundation.

### References

- Alonso, J.M., Stepanova, A.N., Leisse, T.J. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.
- Azpiroz-Leehan, R. and Feldmann, K.A. (1997) T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. *Trends Genet.* **13**, 152–156.
- Balzerque, S., Dubreucq, B., Chauvin, S. *et al.* (2001) Improved PCR-walking for large-scale isolation of plant T-DNA borders. *Bio-techniques*, **30**, 496–504.
- Bechtold, N. and Pelletier, G. (1998) In planta *Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Methods Mol. Biol.* **82**, 259–266.
- Bechtold, N., Ellis, J. and Pelletier, G. (1993) In planta *Agrobacterium*-mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C. R. Acad. Sci. Paris*, **316**, 1194–1199.
- Beetham, P.R., Kipp, P.B., Sawycky, X.L., Arntzen, C.J. and May, G.D. (1999) A tool for functional plant genomics: chimeric RNA/DNA oligonucleotides cause *in vivo* gene-specific mutations. *Proc. Natl Acad. Sci. USA*, **96**, 8774–8778.
- Benfey, P.N., Ren, L. and Chua, N.-H. (1989) The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. *EMBO J.* **8**, 2195–2202.
- Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M. (1989) Genes directing flower development in *Arabidopsis*. *Plant Cell*, **1**, 37–52.
- Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M. (1991) Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development*, **112**, 1–20.
- Brodersen, P., Petersen, M., Pike, H.M., Olszak, B., Skov, S., Odum, N., Jorgensen, L.B., Brown, R.E. and Mundy, J. (2002) Knockout of *Arabidopsis* accelerated-cell-death11 encoding a sphingosine transfer protein causes activation of programmed cell death and defense. *Genes Dev.* **16**, 490–502.
- Brusslan, J.A., Karlin-Neumann, G.A., Huang, L. and Tobin, E.M. (1993) An *Arabidopsis* mutant with a reduced level of cab140 RNA is a result of cosuppression. *Plant Cell*, **5**, 667–677.
- Chen, S., Jin, W., Wang, M., Zhang, F., Zhou, J., Jia, Q., Wu, Y., Liu, F. and Wu, P. (2003) Distribution and characterization of over 1000 T-DNA tags in rice genome. *Plant J.* **36**, 105–113.
- Chory, J., Ecker, J.R., Briggs, S. *et al.* (2000) National Science Foundation-Sponsored Workshop Report: 'The 2010 Project' functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. *Plant Physiol.* **123**, 423–426.
- 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.
- Conklin, P.L., Norris, S.R., Wheeler, G.L., Williams, E.H., Smirnov, N. and Last, R.L. (1999) Genetic evidence for the role of GDP-mannose in plant ascorbic acid (vitamin C) biosynthesis. *Proc. Natl Acad. Sci. USA*, **96**, 4198–4203.
- Dinnyen, J.R., Yadegari, R., Fischer, R.L., Yanofsky, M.F. and Weigel, D. (2004) The role of JAGGED in shaping lateral organs. *Development*, **131**, 1101–1110.
- Edwards, K., Johnstone, C. and Thompson, C. (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucl. Acids Res.* **19**, 1349.
- Eshed, Y., Baum, S.F. and Bowman, J.L. (1999) Distinct mechanisms promote polarity establishment in carpels of *Arabidopsis*. *Cell*, **99**, 199–209.
- Feldmann, K.A. (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *Plant J.* **1**, 71–82.
- Feldmann, K.A., Malmberg, R.L. and Dean, C. (1994) Mutagenesis in *Arabidopsis*. In *Arabidopsis* (Meyerowitz, E. and Somerville, C.R., eds). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 137–172.
- Goff, S.A., Ricke, D., Lan, T.H. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, **296**, 92–100.
- Greene, E.A., Codomo, C.A., Taylor, N.E. *et al.* (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics*, **164**, 731–740.
- Hayashi, H., Czaja, I., Lubenow, H., Schell, J. and Walden, R. (1992) Activation of a plant gene by T-DNA tagging: auxin-independent growth *in vitro*. *Science*, **258**, 1350–1353.
- Ichikawa, T., Nakazawa, M., Kawashima, M. *et al.* (2003) Sequence database of 1172 T-DNA insertion sites in *Arabidopsis* activation-tagging lines that showed phenotypes in T1 generation. *Plant J.* **36**, 421–429.
- Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M. and Last, R.L. (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol.* **129**, 440–450.
- Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R. and Dean, C. (2000) Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science*, **290**, 344–347.
- Kamath, R.S. and Ahringer, J. (2003) Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods*, **30**, 313–321.
- Kamath, R.S., Fraser, A.G., Dong, Y. *et al.* (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, **421**, 231–237.

- 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) Activation tagging of the floral inducer *FT*. *Science*, **286**, 1962–1965.
- Kempin, S.A., Liljgren, S.J., Block, L.M., Rounsley, S.D., Yanofsky, M.F. and Lam, E. (1997) Targeted disruption in *Arabidopsis*. *Nature*, **389**, 802–803.
- 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.
- Konieczny, A. and Ausubel, F.M. (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**, 403–410.
- Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell*, **11**, 2283–2290.
- Li, X., Zhang, Y., Clarke, J.D., Li, Y. and Dong, X. (1999) Identification and cloning of a negative regulator of systemic acquired resistance, SN11, through a screen for suppressors of *npr1-1*. *Cell*, **98**, 329–339.
- Liljgren, S.J., Ditta, G.S., Eshed, Y., Savidge, B., Bowman, J.L. and Yanofsky, M.F. (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature*, **404**, 766–770.
- Lister, C. and Dean, C. (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J.* **4**, 745–750.
- Liu, Y.G. and Whittier, R.F. (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics*, **25**, 674–681.
- Lukowitz, W., Gillmor, C.S. and Scheible, W.R. (2000) Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. *Plant Physiol.* **123**, 795–805.
- Lukowitz, W., Nickle, T.C., Meinke, D.W., Last, R.L., Conklin, P.L. and Somerville, C.R. (2001) *Arabidopsis* *cyt1* mutants are deficient in a mannose-1-phosphate guanylyltransferase and point to a requirement of N-linked glycosylation for cellulose biosynthesis. *Proc. Natl Acad. Sci. USA*, **98**, 2262–2267.
- Marsch-Martinez, N., Greco, R., Van Arkel, G., Herrera-Estrella, L. and Pereira, A. (2002) Activation tagging using the *En-1* maize transposon system in *Arabidopsis*. *Plant Physiol.* **129**, 1544–1556.
- Martienssen, R.A. (1998) Functional genomics: probing plant gene function and expression with transposons. *Proc. Natl Acad. Sci. USA*, **95**, 2021–2026.
- May, S.T., Clements, D. and Bennett, M.J. (2002) Finding your knockout: reverse genetics techniques for plants. *Mol. Biotechnol.* **20**, 209–221.
- McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000) Targeted screening for induced mutations. *Nat. Biotechnol.* **18**, 455–457.
- McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J. and Barton, M.K. (2001) Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. *Nature*, **411**, 709–713.
- Meinke, D.W., Meinke, L.K., Showalter, T.C., Schissel, A.M., Mueller, L.A. and Tzafrir, I. (2003) A sequence-based map of *Arabidopsis* genes with mutant phenotypes. *Plant Physiol.* **131**, 409–418.
- Meyerowitz, E.M. (1987) *Arabidopsis thaliana*. *Annu. Rev. Genet.* **21**, 93–111.
- Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bio-assay with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497.
- Nakazawa, M., Ichikawa, T., Ishikawa, A., Kobayashi, H., Tsuchida, Y., Kawashima, M., Suzuki, K., Muto, S. and Matsui, M. (2003) Activation tagging, a novel tool to dissect the functions of a gene family. *Plant J.* **34**, 741–750.
- Nataraj, A.J., Olivos-Glander, I., Kusakawa, N. and Highsmith, W.E. (1999) Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. *Electrophoresis*, **20**, 1177–1185.
- Neff, M.M., Neff, J.D., Chory, J. and Pepper, A.E. (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant J.* **14**, 387–392.
- Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carlington, J.C. and Weigel, D. (2003) Control of leaf morphogenesis by microRNAs. *Nature*, **425**, 257–263.
- Pan, X., Liu, H., Clarke, J., Jones, J., Bevan, M. and Stein, L. (2003) ATIDB: *Arabidopsis thaliana* insertion database. *Nucl. Acids Res.* **31**, 1245–1251.
- Parinov, S. and Sundaresan, V. (2000) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. *Curr. Opin. Biotechnol.* **11**, 157–161.
- Parinov, S., Sevugan, M., Ye, D., Yang, W.C., Kumaran, M. and Sundaresan, V. (1999) Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in *Arabidopsis*. *Plant Cell*, **11**, 2263–2270.
- Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F. (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature*, **405**, 200–203.
- Petersen, M., Brodersen, P., Naested, H. et al. (2000) *Arabidopsis* map kinase 4 negatively regulates systemic acquired resistance. *Cell*, **103**, 1111–1120.
- Redei, G.P. and Koncz, C. (1992) Classical mutagenesis. In *Methods in Arabidopsis Research* (Koncz, C., Chua, N.-H. and J. Schell, eds). Singapore: World Scientific, pp. 16–82.
- Rhew, R.C., Østergaard, L., Saltzman, E.S. and Yanofsky, M.F. (2003) Genetic control of methyl halide production in *Arabidopsis*. *Curr. Biol.* **13**, 1809–1813.
- Roeder, A.H., Ferrandiz, C. and Yanofsky, M.F. (2003) The role of the REPLUMLESS homeodomain protein in patterning the *Arabidopsis* fruit. *Curr. Biol.* **13**, 1630–1635.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.
- Sessions, A., Burke, E., Presting, G. et al. (2002) A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell*, **14**, 2985–2994.
- Sha, Y., Li, S., Pei, Z., Luo, L., Tian, Y. and He, C. (2003) Generation and flanking sequence analysis of a rice T-DNA tagged population. *Theor. Appl. Genet.* **108**, 306–314.
- 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.
- Sundaresan, 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 Dev.* **9**, 1797–1810.
- Sussman, M.R., Amasino, R.M., Young, J.C., Krysan, P.J. and Austin-Phillips, S. (2000) The *Arabidopsis* knockout facility at the University of Wisconsin-Madison. *Plant Physiol.* **124**, 1465–1467.
- The *Arabidopsis* Genome Initiative (AGI). (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.
- Till, B.J., Reynolds, S.H., Greene, E.A. et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res.* **13**, 524–530.
- Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F. and Komeda, Y. (1996) The *Arabidopsis* *ERECTA*

- gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell*. **8**, 735–746.
- Tzafrir, I., Dickerman, A., Brazhnik, O., Nguyen, Q., McElver, J., Frye, C., Patton, D. and Meinke, D.** (2003) The *Arabidopsis* SeedGenes project. *Nucl. Acids Res.* **31**, 90–93.
- Watson, J.D. and Crick, F.H.** (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
- Weigel, D. and Glazebrook, J.** (2002) *Arabidopsis: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Weigel, D., Ahn, J.H., Blazquez, M.A. et al.** (2000) Activation tagging in *Arabidopsis*. *Plant Physiol.* **122**, 1003–1013.
- Wellmer, F., Riechmann, J.L., Alves-Ferreira, M. and Meyerowitz, E.M.** (2004) Genome-wide analysis of spatial gene expression in *Arabidopsis* flowers. *Plant Cell*. **16**, 1314–1326.
- Wesley, S.V., Helliwell, C.A., Smith, N.A. et al.** (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J.* **27**, 581–590.
- 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*. Isolation of mutations in the cytochrome p450 gene superfamily. *Plant Physiol.* **118**, 743–750.
- Yamada, K., Lim, J., Dale, J.M. et al.** (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science*, **302**, 842–846.
- Zambryski, P.** (1988) Basic processes underlying *Agrobacterium*-mediated DNA transfer to plant cells. *Annu. Rev. Genet.* **22**, 1–30.