

GENOMICS ARTICLE

T-DNA as an Insertional Mutagen in Arabidopsis

Patrick J. Krysan,¹ Jeffery C. Young,^{1,2} and Michael R. Sussman³

Biotechnology Center, University of Wisconsin, 425 Henry Mall, Madison, Wisconsin 53706

INTRODUCTION

Forward genetics begins with a mutant phenotype and asks the question “What is the genotype?” that is, what is the sequence of the mutant gene causing the altered phenotype? Reverse genetics begins with a mutant gene sequence and asks the question “What is the resulting change in phenotype?” These two approaches are fundamentally different, and whereas forward genetics has been in operation for more than a century, the recent avalanche of complete genome sequences has only now created the opportunity for pursuing reverse genetics in an exhaustive and complete manner.

Gene knockouts, or null mutations, are important because they provide a direct route to determining the function of a gene product in situ. Most other approaches to gene function are correlative and do not necessarily prove a causal relationship between gene sequence and function. For example, DNA chips provide an exciting means to discover conditions under which gene expression is regulated on a genomewide scale (DeRisi et al., 1997; Wodicka et al., 1997; Singh-Gasson et al., 1999). However, because factors other than mRNA level alone determine the activity of a gene product in situ, expression studies, even when done on a genomewide scale, cannot prove a causal relationship. By contrast, the availability of a null mutation for the gene of interest allows one to directly monitor the effect this deficiency has on the organism’s ability to function.

There are many ways to implement targeted mutagenesis so as to compromise specific genes. In mice, knockout mutations are now routinely obtained by promoting the homologous recombination of null gene constructs with the genomic wild-type sequence in embryonic stem cells. Provided that the given mutation is not embryonic lethal, “knockout mice” can then be developed in utero by injecting such stem cells into blastocysts (Koller et al., 1989). Yeast

and *Escherichia coli* are other organisms in which homologous recombination is the preferred means for reverse genetics. Although there has been a report of homologous recombination with intact Arabidopsis plants (Kempin et al., 1997), the frequency of this event may be so low as to preclude its use for generating knockout mutations in each of the ~25,000 genes that comprise the 120-Mb genome.

Insertional mutagenesis is an alternative means of disrupting gene function and is based on the insertion of foreign DNA into the gene of interest. In Arabidopsis, this involves the use of either transposable elements (see Parinov et al., 1999, in this issue) or T-DNA. The foreign DNA not only disrupts the expression of the gene into which it is inserted but also acts as a marker for subsequent identification of the mutation. Because Arabidopsis introns are small, and because there is very little intergenic material, the insertion of a piece of T-DNA on the order of 5 to 25 kb in length generally produces a dramatic disruption of gene function. If a large enough population of T-DNA-transformed lines is available, one has a very good chance of finding a plant carrying a T-DNA insert within any gene of interest. Mutations that are homozygous lethal can be maintained in the population in the form of heterozygous plants.

Polymerase chain reaction (PCR) methods have been developed that allow one to easily isolate individual plants that carry a particular T-DNA mutation of interest (McKinney et al., 1995; Krysan et al., 1996). An advantage of using T-DNAs as the insertional mutagen, as opposed to transposons (Martienssen, 1998; Wisman et al., 1998), is that T-DNA insertions will not transpose subsequent to integration within the genome and are therefore chemically and physically stable through multiple generations. The mobility of transposons is not necessarily a bad thing, however. In situations in which multiple members of a gene family are arranged in tandem along a chromosome, the ability of transposons to “hop” to nearby locations provides a convenient method for creating mutations within all of the members of the gene family within a single plant.

Several improvements in Agrobacterium-mediated transformation techniques have made T-DNA a viable method for approaching genomewide mutagenesis. The original root-explant method (Valvekens et al., 1988) allowed one to

¹ These authors contributed equally to this work.

² Current address: Department of Biology, Western Washington University, Bellingham, WA 98225.

³ To whom correspondence should be addressed. E-mail msussman@facstaff.wisc.edu; fax 608-262-6748.

isolate many hundreds and possibly thousands of intact transformed plants, albeit via a laborious tissue culture process. Tens of thousands of transformed plants were beyond reach, however, until Feldmann and Marks (1987) elaborated a method for producing independent T-DNA transgenic lines via seed transformation. The development of transformation methods based on dipping whole plants into *Agrobacterium* suspensions has more recently allowed, in theory, for the production of the hundreds of thousands of insertional mutations necessary for saturation of the genome (Bechtold and Pelletier, 1998; Clough and Bent, 1998).

Saturation of the *Arabidopsis* genome with T-DNA insertions is an experimental goal that requires the actualization of specific quantitative considerations. To date, the quantitative exigencies associated with mutational saturation of the genome by T-DNA have not been fully satisfied. Nevertheless, we have recently established a population of 60,480 T-DNA-transformed lines as a significant step toward the production of genomewide mutations. Access to these lines is now available through the *Arabidopsis* Knockout Facility at the University of Wisconsin (<http://www.biotech.wisc.edu/arabidopsis/default.htm>). This facility will serve the research community by allowing users to screen the entire population of lines for the presence of a T-DNA insert within their gene of interest. The organization of this population of 60,480 lines, as well as the operation of the service facility, is described below.

PROPOSED NOMENCLATURE

The consequences of inserting a T-DNA element into the *Arabidopsis* genome depends on the nature of the T-DNA as well as the precise site of insertion. Figure 1 diagrams several of the possible outcomes of T-DNA insertion and proposes a standard nomenclature for describing them. The "knockon" mutations are a special case in which the T-DNA construct carries a constitutive promoter, such as the cauliflower mosaic virus 35S promoter, capable of driving expression of genes adjacent to the site of insertion (Wilson et al., 1996). The "knockworst" category includes those T-DNA insertion events that lead to large-scale chromosomal rearrangements. Such T-DNA-induced rearrangements have been documented in *Arabidopsis* (Nacry et al., 1998; Laufs et al., 1999).

SATURATING THE GENOME WITH MUTATIONS

Given an infinite number of T-DNA-transformed *Arabidopsis* lines, one should be able to identify a T-DNA insertion within every gene in the genome (with the exception of those genes required for the viability of both the male and female

gametes). It is not practical, however, to generate a population large enough to ensure that every single gene has been mutated. It is therefore important to perform some calculations to estimate how many T-DNA-transformed lines are realistically necessary and sufficient. Three variables determine the probability that a T-DNA insert will be found within a given gene: the size of the gene, the size of the genome, and the number of T-DNA inserts distributed among the population. This relationship is described by the formula shown in the legend of Figure 2A. Of the three independent variables, the only one that is experimentally controllable is the total number of T-DNA inserts implemented within the population.

Figure 2A also shows that the number of T-DNA inserts needed to approach saturation is highly dependent on the

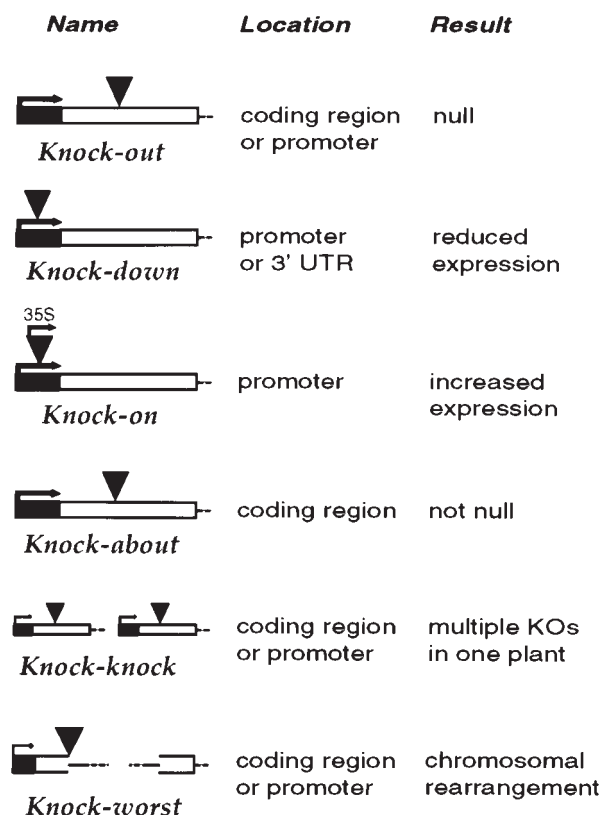


Figure 1. "Knockology."

The insertion of a T-DNA element into an *Arabidopsis* chromosome can lead to many different outcomes. This figure demonstrates several of these possibilities and proposes a standard nomenclature to describe them. The coding region of the gene is shown as a white box, the promoter is the black region with the arrow, and the T-DNA element is represented by a black triangle. KOs, knockouts; UTR, untranslated region.

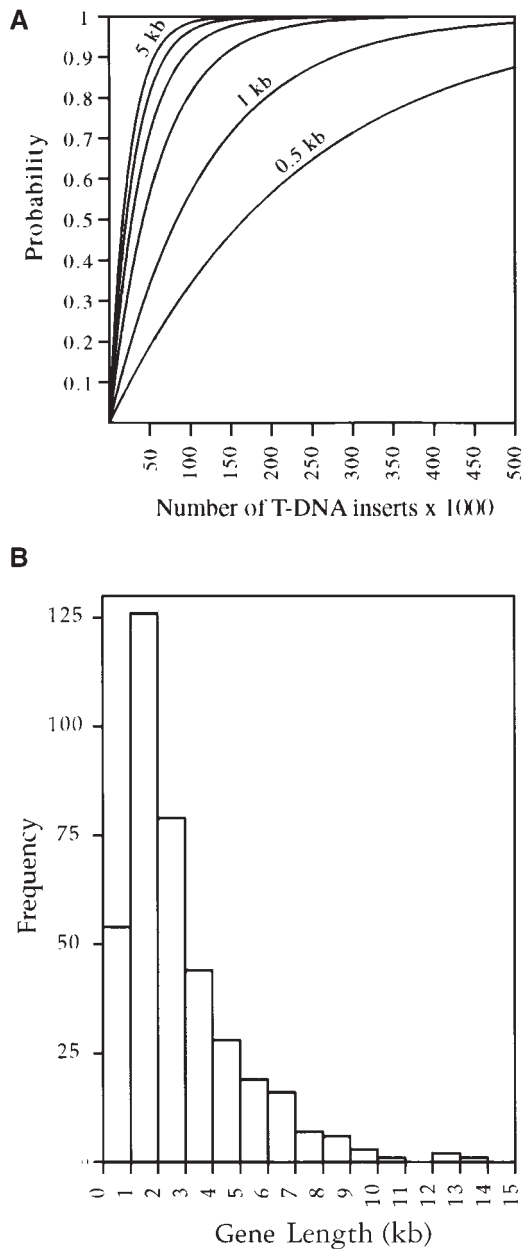


Figure 2. Statistical Profiles of the Arabidopsis Genome.

(A) The probability of finding a T-DNA insertion within a given gene is a function not only of insertional frequency but also of gene length. Curves are drawn for several different gene lengths: 0.5, 1, 2, 3, 4, and 5 kb. The probabilities were calculated using the following formula: $P = 1 - (1 - [x/120,000])^n$, where P = probability of finding one T-DNA insert within a given gene, x = length of the gene in kilobases, and n = number of T-DNA inserts present in the population. This calculation assumes that the haploid Arabidopsis genome is 120 Mb and that T-DNA insertion is random.

(B) The graph presents the frequency distribution of gene sizes in Arabidopsis. We have defined gene to include only that portion of the genomic sequence between the start and stop codons. Three

length of the gene of interest. For example, a 5-kb gene requires 110,000 T-DNA inserts to achieve a 99% probability of being mutated, whereas a 1-kb gene correspondingly necessitates 550,000 T-DNA inserts. It should also be noted that the slope of the curve in Figure 2A flattens out as the probability approaches 100%. Thus, the experimentalist must at some point face the likelihood of diminishing returns when investing time to create additional T-DNA lines.

Because the size of the gene of interest determines the probability of its mutation by T-DNA insertion, we were interested in determining average gene size in Arabidopsis. For this calculation, we defined gene as a genomic DNA sequence, including introns and exons, from which a protein is specified. Sequences upstream and downstream of the sequence flanked by the start and stop codons were not included in our definition. Using this definition of the Arabidopsis gene, we next estimated the size of the productive target region, that is, the portion of the gene within which a T-DNA insertion leads to a null allele. Because T-DNA insertions directly upstream of the start codon would likely lead to null alleles, our omission of upstream regions from our definition of the Arabidopsis gene may result in an underestimate of the actual target size. At the same time, however, it should be considered that insertions at the very end of the coding region may not lead to null alleles; thus, the inclusion of this region within our definition of gene could incur a slight overestimate of target size. In this way, we chose to offset a potential overestimate of target size with a potential underestimate by excluding upstream regions from our definition of gene.

Using published sequence data (Bevan et al., 1998; exclusive of class 6 sequences described therein) and the definition of gene described above, we determined the average length of 386 genes identified within a 1.9-Mb stretch of chromosome 4. A frequency distribution of Arabidopsis gene lengths was generated based on these parameters, as shown in Figure 2B. The median gene length was 2.1 kb, and the mean was 2.8 kb. Bevan et al. (1998) reported a gene density of one gene every 4.8 kb.

Given the median gene length of 2.1 kb determined above, one would require ~280,000 T-DNA inserts to have a 99% chance of mutating a particular gene; a 95% chance would require 180,000 inserts. These numbers provide a framework for determining how many T-DNA-transformed lines need to be created to have a good chance of finding a mutation in the large majority of genes in the Arabidopsis

hundred and eighty-six genes identified within a 1.9-Mb stretch of chromosome 4 (Bevan et al., 1998) were used for this analysis. Frequency indicates the number of individual genes falling within the given length range. Median gene length = 2.1 kb; average gene length = 2.8 kb.

genome. To effectively screen such large populations, efficient and robust protocols must be employed.

SCREENING LARGE POPULATIONS OF T-DNA-TRANSFORMED ARABIDOPSIS LINES

Pool Size

The presence of a T-DNA insertion within any given gene can be easily detected by the proper PCR strategy. Specifically, if PCR is performed using one gene-specific primer and one T-DNA-specific primer, a PCR product is formed only if a T-DNA element has landed either within or very close to the gene of interest. Because PCR is an extremely sensitive method, one can easily screen many thousands of independently transformed *Arabidopsis* plants by means of sample pooling (McKinney et al., 1995; Krysan et al., 1996). Given the requirement for screening hundreds of thousands of T-DNA transformed lines (see above), any pooling strategy will have to take the upper limit of pool size into consideration.

In early experiments, DNA was extracted from groups of 10 lines and then combined to make fifty-three pools, each pool representing DNA from 100 lines (McKinney et al., 1995). No further pooling was performed, and thus fifty-three PCR reactions were required to search the population of 5300 transformed lines. The high number of PCR reactions in these experiments was offset in part by the use of degenerate primers designed to identify several closely related genes. Krysan et al. (1996) demonstrated that pool sizes of >1000 lines could be searched using gene-specific primers. In this case, starting with a population of 9100 T-DNA-transformed lines, DNA was extracted from pools of 100 lines. DNA from thirteen such extractions was then combined to make superpools, each of which represented the DNA of 1300 lines. Such large pool sizes lowered the number of PCRs required for screening the 9100 lines to seven. This was accomplished in part by using a stringent PCR annealing temperature (65°C) and long primers (29 bp) coupled with the use of a highly processive *taq* polymerase (Krysan et al., 1996).

To further probe the upper size limit on DNA pools, we performed a PCR experiment using various amounts of template DNA derived from a pool of 225 independently transformed lines. These PCRs used the T-DNA left border primer and one gene-specific primer to detect a T-DNA insert known to be present in this pool of 225. As shown in Table 1, three different amounts of template DNA were tested, using sixteen replicates of each PCR to determine the reproducibility with which the T-DNA insert could be detected. In this analysis, it was concluded that ~208 copies of the T-DNA insert are required in a complex pool of DNA to ensure that the insertion is reliably detected. It was also found that PCR efficiency is attenuated when >125 ng of pooled

DNA from these T-DNA lines is used per PCR. These limitations therefore set the maximum useful pool size under our conditions at ~2350 lines per pool.

Pool Architecture

The manner in which a large population of T-DNA-transformed lines is organized for pooling requires care to ensure the efficient establishment of a useful resource. Pool architectures such as the multiplex approach suggested by Azpiroz-Leehan and Feldmann (1997) are designed to provide a reliable degree of systematization. Such pooling schemes, however, require significant set-up time and resources.

We have recently organized a population of 60,480 T-DNA-transformed lines by using the simple strategy outlined in Figure 3A, which had proven successful on smaller populations (Krysan et al., 1996). This collection of lines was created by Dr. Rick Amasino and colleagues using *Arabidopsis thaliana* ecotype Wassilewskija transformed with a derivative of the pD991 T-DNA vector, which happens to carry a small portion of the *APETALA* promoter. pD991 itself is a derivative of the pCGN1547 binary vector (McBride and Summerfelt, 1990). Kanamycin-resistant seedlings (T_1 generation) were isolated and transferred to soil at a density of nine seedlings per pot. Because physical space and human resources were limiting, T_2 generation seeds were collected in bulk from each pot containing nine plants. These batches of seed are called "pools of nine" because each represents seed collected from nine independently transformed parent plants. Rather than extract DNA individually from each of the 6720 pools of nine, we chose to first consolidate the collection into a manageable number of samples. We began this process by creating "pools of 225." A pool of 225 is a batch of seed that is derived from 225 independently transformed parent plants. These pools of 225 were created by scooping equal portions of seed from 25 separate pools of nine into a single container. In this manner, the entire collection of

Table 1. The Limits of Sensitivity for Detecting a Specific T-DNA Insert

Total Template DNA ^a	Copy Number of T-DNA ^b	Success Rate ^c
12 ng	208 copies	16/16
4 ng	69 copies	11/16
1 ng	17 copies	8/16

^a Indicates the total amount of template DNA present in each PCR reaction.

^b Indicates the number of copies of this particular T-DNA insertion that should be present in each PCR reaction.

^c Indicates how many of the 16 PCR reactions performed were able to detect the T-DNA insert.

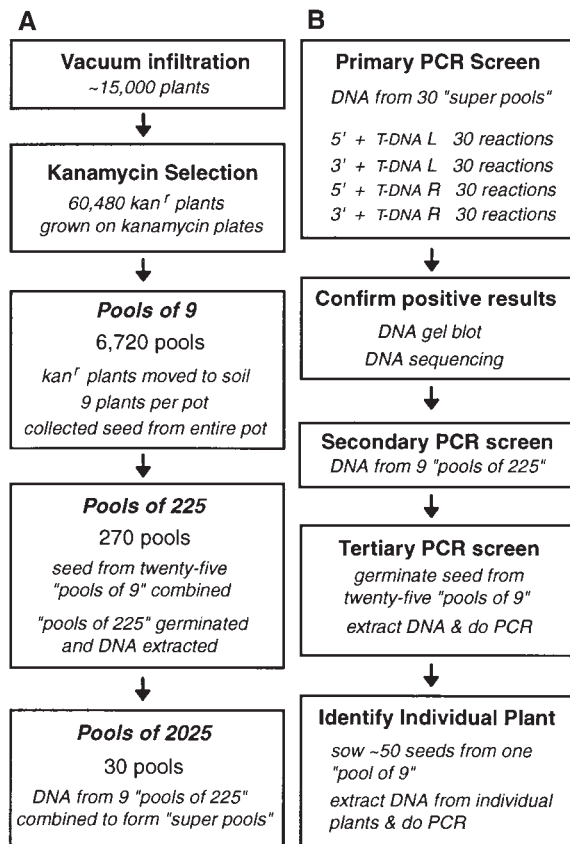


Figure 3. Organization and Screening of 60,480 T-DNA-Transformed Arabidopsis Lines.

(A) Pooling strategy.

(B) Insertion screening strategy.

5' and 3' refer to PCR primers specific for the gene of interest. T-DNA L and T-DNA R refer to PCR primers specific for the T-DNA border regions. kan^r, kanamycin resistant.

60,480 plants was reduced to an ordered collection of 270 pools of 225. As an example of the time required to handle these quantities, it takes approximately one person-month to aliquot 6720 pools of nine into 270 pools of 225.

Seed from each of the 270 unique pools of 225 was then germinated in a liquid culture, and genomic DNA was extracted from the resulting seedling pools. This work resulted in the generation of 270 separate DNA samples, with each DNA sample representing 225 independently transformed plants. Finally, DNA superpools were formed in which each superpool contained the DNA extracted from nine separate pools of 225. In this manner, 30 superpools were created, with each DNA superpool representing 2025 (225 × 9) independently transformed lines. The entire population of 60,480 transformed plants is represented within these 30 DNA superpools.

This ordered population of 60,480 T-DNA insertion lines can then be exhaustively screened with 120 PCRs. However, we generally limit our initial screens to 60 reactions by using only the T-DNA left border primer. Preliminary results indicate that the T-DNA left border is detected two to three times more often than the right border in this population. A predominance of intact left border sequences has been previously documented (Castle et al., 1993; Krysan et al., 1996). Once a T-DNA insertion line is identified in a superpool, the process of isolating a single plant requires the short series of experiments shown in Figure 3B.

THE IMPORTANCE OF TESTING EMPIRICALLY PCR PRIMERS

T-DNA Border Primers

Our experience with screening pooled T-DNA samples has revealed that the choice of PCR primers for the T-DNA border region is critical to the optimization of the procedure. Most importantly, we have found that there is currently no good method for predicting the suitability of a particular primer for use as a T-DNA border primer in a knockout screen. It is therefore necessary to experimentally test a number of candidate T-DNA border primers within the context of the actual DNA superpools. This process should ultimately allow one to identify a T-DNA border primer that consistently amplifies a known T-DNA insertion with a minimum of artifacts. In practice, we have found that all of the T-DNA border primers we have tested produce some artifactual bands, probably due to T-DNA rearrangements and concatamers. Nevertheless, by systematically testing several different T-DNA border primers, we were able to identify suitable primers specific for the particular T-DNA border sequences present in the pD991 vector used in the creation of our 60,480 lines.

Gene-Specific Primer Design

Primers that appear to work under standard PCR conditions often fail to give the necessary sensitivity to detect a rare T-DNA insert within a population of pooled T-DNA insertion lines. It is therefore necessary to test the primers under the same PCR conditions that are used to identify T-DNA insertions. We typically design our primers according to a few simple guidelines. In particular, we utilize primers 29 base pairs in length that have a G+C content between 34 and 50%. We avoid a G+C content of >50% in positions 19 to 29 of the primer, and we allow G+C to equal 0 or 1 at positions 28 and 29 (3' end) of the primer. These guidelines do not substitute for empirical tests of primer performance. In addition, it is necessary to test the gene-specific primers in

combination with the T-DNA border primers to ensure compatibility.

CHARACTERIZATION OF PHENOTYPES

The identification of knockout mutants is the first step toward describing the function of a gene. After the isolation of a mutant line, plants homozygous for the mutation must be identified, outcrossed, and analyzed to ensure that only one T-DNA insertion is present. With a confirmed mutant in hand, the next step is to determine the consequences of the mutation on growth and development relative to the wild type. However, it has become apparent that many knockout mutants have no readily identifiable phenotype. For example, of the 17 mutants described in Krysan et al. (1996), none displays an altered phenotype unless grown under specific conditions (Hirsch et al., 1998; P.J. Krysan, J.C. Young, and M.R. Sussman, unpublished results). The flow chart shown in Figure 4 demonstrates the many steps that are necessary when one is analyzing the phenotypic consequences of a particular knockout mutation.

Functional redundancy among the members of a gene family is a likely reason for the frequently observed lack of an identifiable phenotype associated with knockout muta-

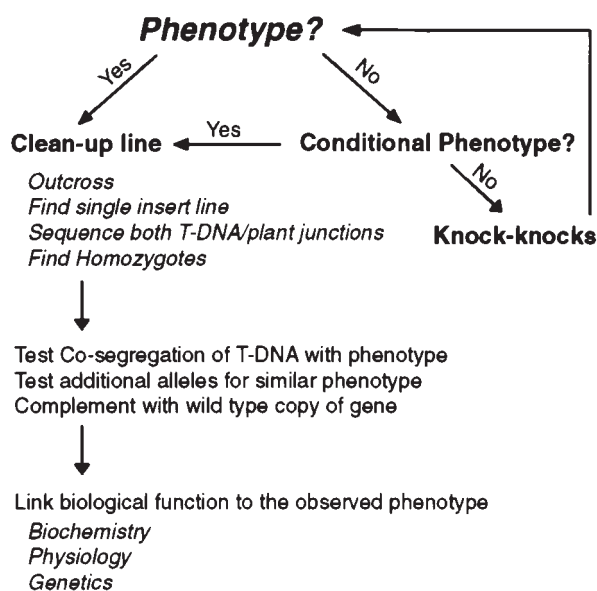


Figure 4. Using Insertional Mutations to Understand Biological Function.

The flow chart outlines the steps for characterizing the phenotypic consequences of a particular T-DNA-induced mutation. "Knock-knocks" refers to the process of genetic crossing to create plants that carry mutations in multiple members of a given gene family (see Figure 1).

tions (Hua and Meyerowitz, 1998). To test for functional redundancy, genetic crosses between plants that bear mutations in different members of the gene family can be performed, resulting in the formation of "knock-knock" mutations (see Figure 1). A directed approach can be taken in which knock-knocks involving gene family members with the highest sequence similarity, or with similar expression patterns, are created. As the number of Arabidopsis T-DNA insertion lines increases, it will become possible to obtain knockout mutations of most if not all members of a gene family, thereby making it possible to test all mutant combinations. However, the number of possible multiple-mutant genotypes (as described by the formula 2^n , where n is the number of gene family members) becomes daunting when one considers gene families with >10 members. However, an ordered and tractable approach is available.

First, double-mutant lines are created and, when viable, crossed to lines homozygous for a mutation of a third member of a given gene family. Because the T-DNA insert associated with each mutation serves as a marker, PCR genotyping of complex mutant backgrounds is possible. In this cumulative manner, any number of the members in a given gene family can theoretically be knocked out within a single line, provided the genes are not too closely linked on the chromosome. Once a variety of multiple-mutant lines is available, the lines can in turn be crossed with each other to obtain complex segregating populations. At any point, the genotype for seedlings with interesting phenotypes in these segregating populations can be determined using PCR.

Another possible reason for the lack of observable phenotypes is that individual gene family members may have evolved to function only under specific physiological conditions. Thus, unless the mutant plant is placed under a condition in which the target gene is required, no phenotype is observed (Hirsch et al., 1998). To test for conditional phenotypes in a gene of unknown function, one must employ a broad panel of physiologically meaningful conditions. Ultimately, the ability to assay the expression of a large number of genes simultaneously will assist in determining the effect of knockout mutations on plant growth and development. Progress in DNA chip microarray technology (Singh-Gasson et al., 1999), in combination with Arabidopsis sequence information, will make genome-wide expression studies possible in the near future.

CONFIRMING CORRELATIONS BETWEEN MUTATION AND PHENOTYPE

The goal of reverse genetics is the identification of a phenotype that is caused by mutation of a particular gene. Once such a phenotype has been observed, several steps must be taken to prove that the phenotypic characteristic is indeed controlled by the gene of interest. The first step is to follow the segregation of the T-DNA over multiple genera-

tions and score the corresponding phenotypes. One can easily determine the precise genotype of large numbers of individual plants by using the T-DNA insertion as a PCR marker for the mutant locus. Such analysis, however, does not prove that the PCR-identified, T-DNA-induced mutation is responsible for the phenotype rather than a closely linked, unrelated mutation.

To prove definitively that the insertional mutation causes the phenotype, one must either isolate additional mutant alleles for the locus or complement the mutation by introducing a wild-type copy of the gene into mutant plants by using transgenic technology. If additional mutant alleles are available, they will provide the quickest route to confirming or refuting the role played by the insertionally mutated gene in controlling the observed phenotype. If the same phenotype is found to be linked to the same T-DNA insertion in several independently transformed *Arabidopsis* plants, one could make a strong argument that the mutation is indeed causing the phenotype.

One of the benefits of generating a large collection of T-DNA-transformed lines is that the probability of finding more than one T-DNA insert in a given gene is quite high. For instance, given a 95% chance of finding a single insert in a particular gene, one would have a 90% chance of finding two independent inserts in that same gene and an 86% chance of finding three alleles. Having access to a large population of T-DNA-transformed lines could thus supplant the labor-intensive process of transgenic complementation with the more efficient process of multiple allele analysis.

ESTABLISHMENT OF A NATIONAL ARABIDOPSIS KNOCKOUT FACILITY

We have recently established a service facility at the University of Wisconsin that will provide the *Arabidopsis* research community with access to our population of 60,480 T-DNA-transformed lines. Detailed information about the operation of the facility can be found by visiting its Web site (<http://www.biotech.wisc.edu/arabidopsis/default.htm>).

Using gene-specific primers provided by the user, our facility will perform PCRs that screen the entire population of 60,480 lines for the presence of a T-DNA insert within the gene of interest. The resulting PCRs will be mailed to the user, who will be responsible for analyzing the reactions by gel electrophoresis, DNA gel blotting, and DNA sequencing to determine if any knockouts of the user's gene are present in the population. If a positive result is obtained in the first round of PCR, the user can then request that a second round of PCR be performed by the facility, whereby the particular pools of 225 that carries the knockout of interest will be identified. Finally, the knockout facility will supply the user with seed from the 25 pools of nine that correspond to the pool of 225. The user will then perform DNA extractions

and PCR to determine which pool of nine contains the user's mutant and will ultimately isolate the individual mutant plant.

BEYOND PCR SCREENING

The process of PCR screening for individual knockout mutations is an efficient and fruitful approach to reverse genetics. This strategy allows one to focus resources and energies on a small number of interesting genes. As the age of high-throughput genomics arrives, it is apparent that one could also pursue an alternative strategy. Rather than searching for mutations in particular genes, one could simply begin cataloging the locations of all of the T-DNA inserts present in the entire population (Bouchez and Höfte, 1998). Methodologies such as plasmid rescue, inverse PCR, and thermal asymmetric interlaced PCR (Allen et al., 1994; Liu et al., 1995) all provide effective strategies for isolating the genomic DNA sequence immediately adjacent to the site of individual T-DNA integration.

The *Arabidopsis* Knockout Facility at the University of Wisconsin will soon begin a program of isolating and sequencing the DNA that flanks the T-DNA inserts present in its population of 60,480 lines. This strategy will allow us to characterize most, if not all, of the T-DNA inserts present in each pool of nine lines. A computer database will then be established in which all of the T-DNA flanking sequences will be stored, along with a notation to indicate the corresponding pool of nine lines. This database can then be searched for the presence of flanking sequences homologous to any gene of interest, and the corresponding pool of nine can be ordered directly from the stock center, eliminating the need for large-scale PCR screens.

CONCLUSION

With three-quarters of the *Arabidopsis* genome already sequenced and the expected completion of the entire genome within the next year, the era of reverse genetics should yield simple and direct routes for exploring gene function. In conjunction with other emerging genomic technologies, reverse genetic analysis will provide a solid foundation upon which to build a more complete understanding of the complex interactions among the thousands of different genes present in *Arabidopsis*.

ACKNOWLEDGMENTS

The authors thank Dr. Rick Amasino and his laboratory and Sandra Austin-Phillips for producing the T-DNA-tagged lines; they also thank Heather Burch and Sarah Graham for growing tissue and

extracting DNA. Thanks also to Pete Jester, Laura Katers, and Sean Monson for technical assistance. This work was supported by Grant No. DBI 9872638 from the National Science Foundation.

Received August 16, 1999; accepted October 27, 1999.

REFERENCES

- Allen, M.J., Collick, A., and Jeffreys, A.J. (1994). Use of vectorette and subvectorette PCR to isolate transgene flanking DNA. *PCR Methods Appl.* **4**, 71–75.
- Azpiroz-Leehan, R., and Feldmann, K.A. (1997). T-DNA insertion mutagenesis in *Arabidopsis*: Going back and forth. *Trends Genet.* **13**, 152–156.
- 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.
- Bevan, M., et al. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391**, 485–488.
- 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.
- DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686.
- Feldmann, K.A., and Marks, M.D. (1987). *Agrobacterium* mediated transformation of germinating seeds of *Arabidopsis thaliana*: A non-tissue culture approach. *Mol. Gen. Genet.* **208**, 1–9.
- Hua, J., and Meyerowitz, E.M. (1998). Ethylene responses are negatively regulated by a receptor gene family in *Arabidopsis thaliana*. *Cell* **94**, 261–271.
- Hirsch, R.E., Lewis, B.D., Spalding, E.P., and Sussman, M.R. (1998). A role for AKT1 potassium channel in plant nutrition. *Science* **280**, 918–921.
- Kempin, S.A., Liljegren, S.J., Block, L.M., Rounsley, S.D., Lam, E., and Yanofsky, M.F. (1997). Inactivation of the *Arabidopsis* AGL5 MADS-box gene by homologous recombination. *Nature* **389**, 802–803.
- Koller, B.H., Hagemann, L.J., Doetschman, T., Hageman, J.R., Huang, S., Williams, P.J., First, N.L., Maeda, N., and Smithies, O. (1989). Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **86**, 8927–8931.
- 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.
- Laufs, P., Autran, D., and Traas, J. (1999). A chromosomal paracentric inversion associated with T-DNA integration in *Arabidopsis*. *Plant J.* **18**, 131–139.
- 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.
- Martienssen, R.A. (1998). Functional genomics: Probing plant gene function and expression with transposons. *Proc. Natl. Acad. Sci. USA* **95**, 2021–2026.
- McBride, K.E., and Summerfelt, K.R. (1990). Improved binary vectors for *Agrobacterium* mediated plant transformation. *Plant Mol. Biol.* **14**, 269–276.
- 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.
- Nacry, P., Camilleri, C., Courtial, B., Caboche, M., and Bouchez, D. (1998). Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics* **149**, 641–650.
- Parinov, S., Sevugan, M., Ye, D., Yang, W.-C., Kumaran, M., and Sundaresan, V. (1999). Analysis of Flanking Sequences from *Disso-ciation* Insertion Lines: A database for reverse genetics in *Arabidopsis*. *Plant Cell* **11**, 2263–2270.
- Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micro-mirror array. *Nat. Biotechnol.* **17**, 974–978.
- Valvekens, D., Van Montagu, M., and Van Lijsebettens, M. (1988). *Agrobacterium tumefaciens* mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. *Proc. Natl. Acad. Sci. USA* **85**, 5536–5540.
- Wilson, K., Long, D., Swinburne, J., and Coupland, G. (1996). A Dissociation insertion causes a semidominant mutation that increases expression of *TINY*, an *Arabidopsis* gene related to *APETALA2*. *Plant Cell* **8**, 659–671.
- Wisman, E., Hartmann, U., Sagasser, M., Baumann, E., Palme, K., Hahlbrock, K., Saedler, H., and Weisshaar, B. (1998). Knock-out mutants from an En-1 mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proc. Natl. Acad. Sci. USA* **95**, 12432–12437.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., and Lockhart, D.J. (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**, 1359–1367.