Targeting Induced Local Lesions IN Genomes (TILLING) for Plant Functional Genomics

Claire M. McCallum, Luca Comai, Elizabeth A. Greene, and Steven Henikoff*

Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109–1024 (C.M.M., E.A.G., S.H.); Molecular and Cellular Biology Program (C.M.M.) and Department of Botany (L.C.), University of Washington, Seattle, Washington 98195; and Howard Hughes Medical Institute Research Laboratories, Seattle, Washington 98109 (S.H.)

One of the most important breakthroughs in the history of genetics was the discovery that mutations can be induced (Muller, 1930; Stadler, 1932). The high frequency with which ionizing radiation and certain chemicals can cause genes to mutate made it possible to perform genetic studies that were not feasible when only spontaneous mutations were available. As a result, much of our understanding of genetics of higher organisms is based upon studies utilizing induced mutations for analyzing gene function. Alkylating agents, which yield predominantly point mutations, have been especially valuable, since the resulting altered and truncated protein products help to precisely map gene and protein function. Because of the high mutational density and the great utility of point mutations, traditional chemical mutagenesis methods have continued to be popular in phenotypic screens despite the development of other mutagenic tools such as transposon mobilization (Bingham et al., 1981).

With the recent expansion of sequence databanks, locus-to-phenotype reverse genetic strategies have become an increasingly popular alternative to phenotypic screens for functional analysis. Sequence information alone may be sufficient to consider a gene to be of interest, because sequence comparison tools that detect protein sequence similarity to previously studied genes often allow a related function to be inferred. Hypotheses concerning gene function that are generated in this way must be confirmed empirically. Experimental determination of gene function is desirable in other situations as well, for example, when a genetic interval has been associated with a phenotype of interest. In such cases, the functions of genes in an interval can be inferred by using reverse genetic methods. Routine reverse genetics (Scherer and Davis, 1979) has been an important factor in the popularity of baker's yeast over the past two decades, and the RNAi technique (Fire et al., 1998) now provides *Caenorhabditis elegans* investigators with a routine knockout method that has enjoyed huge popularity over the past year (Sharp, 1999). In most other eukaryotes, however, the situation remains unsatisfactory.

In plants, the two most common methods for producing reduction-of-function mutations are antisense RNA suppression (Schuch, 1991; de Lange et al., 1995; Hamilton et al., 1995; Finnegan et al., 1996) and insertional mutagenesis (Altmann et al., 1995; Smith et al., 1996; Azpiroz-Leehan and Feldmann, 1997; Long and Coupland, 1998; Martienssen, 1998; Pereira and Aarts, 1998; van Houwelingen et al., 1998; Speulman et al., 1999). However, antisense RNA suppression requires considerable effort for any given target gene before knowing whether it will work, and insertional mutagenesis occurs at a low frequency per genome. There is current interest in RNAi-related suppression (Waterhouse et al., 1998). However, its efficacy is not yet clear; for example, epigenetic phenotypes can be variegated and unpredictable (Que and Jorgensen, 1998). Because these techniques rely either on *Agrobacterium* T-DNA vectors for transmission or on an endogenous tagging system, their usefulness as general reverse genetics methods is limited to very few plant species. Moreover, these techniques produce a very limited range of allele types. Therefore, as the amount of sequence data grows for Arabidopsis and other organisms, it is important to develop genome-scale reverse genetic strategies that are automated, broadly applicable, and capable of creating the wide range of mutant alleles that is needed for functional analysis.

We have introduced a new reverse genetic strategy that combines the high density of point mutations provided by traditional chemical mutagenesis with rapid mutational screening to discover induced lesions (McCallum et al., 2000). TILLING (Targeting Induced Local Lesions IN Genomes) combines chemical mutagenesis (Koornneef et al., 1982) with a sensitive mutation detection instrument. In a pilot experiment, DNA from a collection of EMS-mutagenized Arabidopsis plants was pooled, subjected to PCR amplification, and screened for mutations using denaturing HPLC (DHPLC). DHPLC detects mismatches in heteroduplexes created by melting and annealing of heteroallelic DNA. Among the lesions detected were base transitions causing missense and nonsense changes that can be used for phenotypic analyses.

* Corresponding author; e-mail shenikof@fhcrc.org; fax 206–667–5889.

TILLING is suitable for any organism that can be heavily mutagenized, even those that lack genetic tools. Starting with a homozygous population is desirable, because DHPLC will detect polymorphisms. Nevertheless, this strategy can be applied to species and hybrids that cannot be practically homozygosed: we and others have detected rare polymorphisms in a heteroallelic background using DHPLC (C.M. McCallum and S. Henikoff, unpublished data; N. Suter and E. Ostrander, personal communication). The general applicability of TILLING makes it appropriate for genetic modification of crops, and there may be agricultural interest in producing phenotypic variants without introducing foreign DNA of any type into a plant's genome.

The strategy is illustrated in Figure 1. The steps are: (a) EMS mutagenesis (Redei and Koncz, 1992; Feldmann et al., 1994; Lightner and Caspar, 1998); (b) DNA preparation and pooling of individuals; (c) PCR amplification of a region of interest; (d) denaturation and annealing to allow formation of heteroduplexes; (e) DHPLC, where the presence of a heteroduplex in a pool is detected as an extra peak in the chromatogram; (f) identification of the mutant individual; and (g) sequencing of the mutant PCR product.

An advantage of TILLING is that the likelihood of recovering a deleterious mutation can be calculated in advance. A calculation is possible, since EMS produces primarily C/G to T/A transitions (Ashburner, 1990). For example, 20 of 23 LEAFY EMS-generated mutations are from C to T, resulting in C/G to T/A transitions (http://www.salk.edu/LABS/pbio-w/lfyseq.html). The probability of discovering deleterious alleles can be maximized by judicious choice of the region to be TILLed (Fig. 2). Furthermore, by choosing coding regions that are evolutionarily conserved, it becomes more likely that missense mutations with detrimental effects on gene function will be obtained. Splice junction mutations are also potentially deleterious.

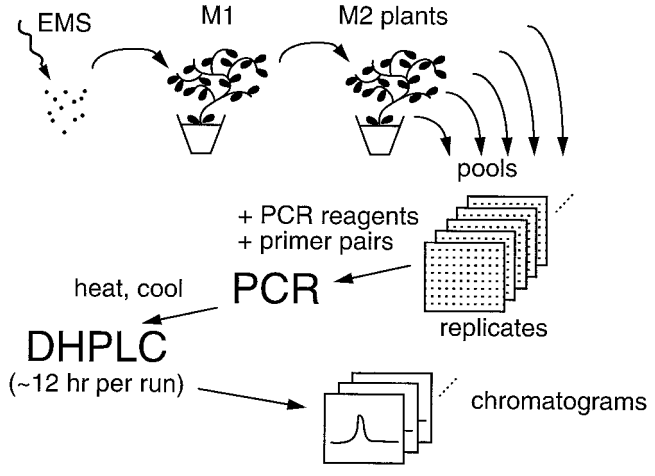

Figure 1. Schematic depicting the TILLING strategy applied to a plant such as Arabidopsis.

Although TILLING minimizes the effort required to find mutations, ascertaining the resulting phenotype requires further work. Chemical mutagenesis introduces background mutations that can make phenotypic analysis uncertain, and multiple generations of outcrossing may be desirable. However, a rapid strategy is available if two independent deleterious lesions are found: the two individuals can be crossed and their progeny genotyped by DHPLC. A phenotype attributable to the two non-complementing mutations will be found in every heteroallelic individual, whereas non-complementing background mutations will assort independently.

TILLING is appropriate for both small- and large-scale screening, because the high density of mutations requires relatively few individual plants. Even for genome-wide TILLING, our pilot screen data suggest to us that a collection of 10,000 reference Arabidopsis plants will suffice for obtaining the desired mutations from just a single primer pair per gene. By using multiple primer pairs to scan a gene of average size (or larger), fewer plants are needed. For example, fewer than 1,000 plants were used in our pilot study. With our current protocol, operation of a single DHPLC machine is expected to discover at least one knockout lesion per amplified gene in 1 to 2 weeks (1,000–2,000 runs, yielding 10–20 lesions, of which 5% will be stop codons). Thus, TILLING is an attractive strategy for a small research group.

TILLING might also be envisioned on a large scale. Unlike clone-based reverse genetic methods, TILLING utilizes rapidly advancing technology (such as DHPLC) that is being developed for high-throughput polymorphism detection. Even with current technology, it should be possible to increase the size of pools over what was used in our pilot project by sacrificing some level of sensitivity, which only marginally reduces throughput. Another way to increase throughput is to use higher doses of EMS than was used in the pilot screen (Koornneef et al., 1982; Sega, 1984; Schy and Plewa, 1989), and we estimate that this would double the rate of mutation.

Most steps of TILLING are suitable for automation. The choice of PCR amplicon can be automated (for high-throughput) and streamlined for interactive use (by users requesting genes for TILLING). By assigning a score to regions of target genes based on the likelihood of obtaining desirable mutations (Fig. 2), genes and gene regions can be rank-ordered, and the ranks can be used for primer selection. Data analysis can also be automated. Two classes of data are generated: DHPLC chromatograms and sequence traces. Software for reading chromatograms does not yet exist; however, software for reading sequence traces from heterozygotes is available (Nickerson et al., 1997).

DHPLC is only one of the promising technologies being developed for polymorphism discovery that can be applied to TILLING. One possible future di-

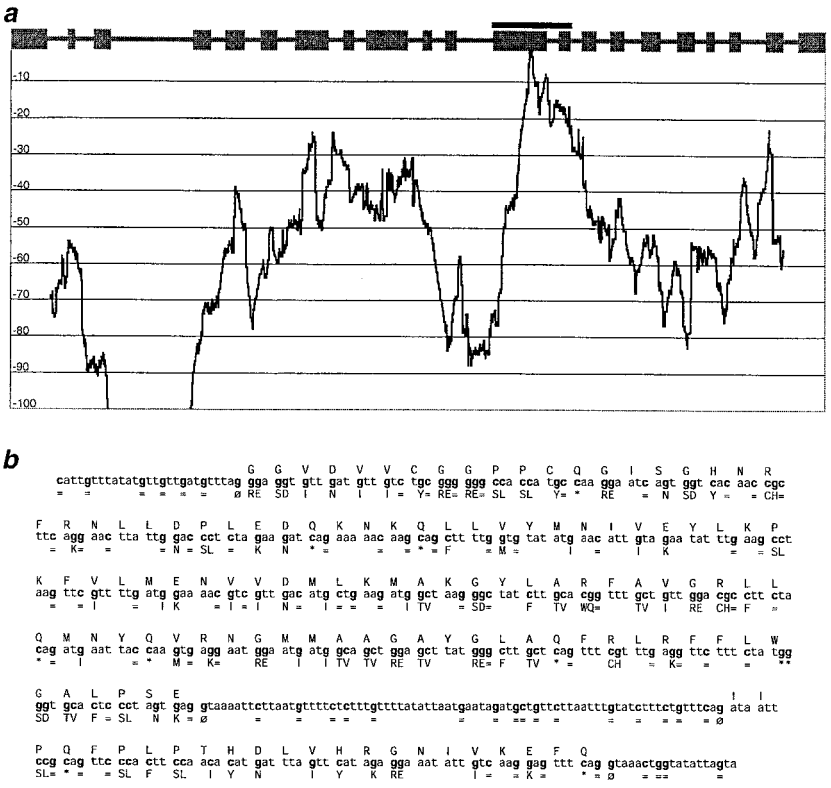

Figure 2. Optimal amplicon selection for deleterious mutation discovery. a, Graphical representation of the relative susceptibility of each 500-bp amplicon to C/G to T/A transitions causing a deleterious mutation in the Arabidopsis CMT3 gene. Exons are indicated as shaded boxes above the plot. Each point on the plot is the sum of scores calculated for a 500-bp amplicon window centered at that residue. A residue susceptible to a nonsense change scored +6, to a missense change scored 0, to a silent change scored −1, and to a splice junction mutation scored +4. The scoring system is based on the potential overall frequency of these changes in genes of Arabidopsis. Bar delimits region analyzed in b. b, Sites that are susceptible to the C/G to T/A transition mutations are indicated under the DNA sequence for the amplicon. Each amino acid of the coding sequence is indicated above its codon. The consequence for each mutation is indicated below. Letters indicate a missense change, = indicates a silent change, * indicates a nonsense change, and ϕ indicates a splice site mutation.

rection is the use of mismatch repair enzymology to detect heteroduplexes, one example being the CEL I endonuclease from celery (Oleykowski et al., 1998). CEL I recognizes a mismatch and cleaves exactly at the 3′ side of the mismatch. Cutting by CEL I followed by denaturing gel electrophoresis can pinpoint the precise base position of a mismatch. Once the location of a mismatch is determined, the base change can be inferred, since EMS produces mainly C/G to T/A transition mutations. Therefore, the CEL I assay would not only limit the need for sequencing during discovery, but would also reduce the amount of effort required to identify the individuals that have the desirable mutations. Another direction is to increase sample pooling to exploit continuing improvements in the detection and resolution of rare DNA molecules within mixtures. Improved detection methods include those that utilize capillary electrophoresis: constant denaturant capillary electrophoresis and single-stranded conformational polymorphism (Larsen et al., 1999; Li-Sucholeiki et al., 1999; Nataraj et al., 1999). Capillary electrophoresis has been successfully exploited for high-throughput DNA sequencing (Kheterpal and Mathies, 1999), and we anticipate its adaptation for high-throughput polymorphism detection.

We are currently establishing a collection of approximately 10,000 mutagenized reference M2 Arabidopsis plants for large-scale TILLING, which could become a community-wide resource. We envision that someone interested in using a TILLING resource will be seeking mutations in a sequenced gene of interest. This greatly simplifies the task of database maintenance, because all that is needed to find mutations is to perform a similarity search using the sequence of interest to query the database of mutant sequences. The mutation itself will be easily pinpointed as (presumably) the only non-matching alignment pair. Reference plants will be made available from the Arabidopsis Biological Resource Center (Ohio State University, Columbus).

Plants are especially well suited to our strategy, because they can be self-fertilized and seeds can be easily stored. This does not mean that TILLING is just for plants: mouse ES cells can be mutagenized with EMS and stored frozen (Schimenti and Bucan, 1998) and so development of strategies for plants might prove to facilitate high-throughput technology for mammalian systems, which counters the perception that plant biotechnology borrows from animal systems. Arabidopsis is the obvious choice for the prototypic implementation of high-throughput TILLING, because it is the only plant species with a nearly complete gene sequence database. The greatest utility of TILLING might be for crop and other model plants such as rice and *Medicago truncatula*, which are currently being subjected to large-scale genome and cDNA analyses. Sequence data provided by these efforts provide fodder for TILLING, a reverse genetic strategy that does not require advanced genetic tools.

Received February 14, 2000; accepted February 22, 2000.

LITERATURE CITED

Altmann T, Felix G, Jessop A, Kauschmann A, Uwer U, Peña-Cortés H, Willmitzer L (1995) Mol Gen Genet **247:** 646–652

Ashburner M (1990) *Drosophila*, A Laboratory Handbook. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Azpiroz-Leehan R, Feldmann KA (1997) Trends Genet **13:** 152–156

Bingham PM, Levis R, Rubin GM (1981) Cell **25:** 693–704

de Lange P, van Blokland R, Kooter JM, Mol JN (1995) Curr Top Microbiol Immunol **197:** 57–75

Feldmann KA, Malmber RL, Dean C (1994) *In* EM Meyerowitz, CR Somerville, eds, Arabidopsis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 137–172

Finnegan EJ, Peacock WJ, Dennis ES (1996) Proc Natl Acad Sci USA **93:** 8449–8454

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Nature **391:** 806–811

Hamilton AJ, Fray RG, Grierson D (1995) Curr Top Microbiol Immunol **197:** 77–89

Kheterpal I, Mathies RA (1999) Anal Chem **71:** 31A–37A

Koornneef M, Dellaert LW, van der Veen JH (1982) Mutat Res **93:** 109–123

Larsen LA, Christiansen M, Vuust J, Anderson PS (1999) Hum Mutat **13:** 318–327

Lightner J, Caspar T (1998) *In* J Martinez-Zapater, J Salinas, eds, Methods on Molecular Biology, Vol 82. Humana Press, Totowa, NJ, pp 91–104

Li-Sucholeiki XC, Krhapko K, Andre PC, Marcelino LA, Karger BL, Thilly WG (1999) Electrophoresis **20:** 1224–1232

Long D, Coupland G (1998) Methods Mol Biol **82:** 315–328

Martienssen RA (1998) Proc Natl Acad Sci USA **95:** 2021–2026

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Nat Biotechnol **18:** 455–457

Muller HJ (1930) J Genet **22:** 299–334

Nataraj AJ, Olivos-Glander I, Kusukawa N, Highsmith WEJ (1999) Electrophoresis **20:** 1177–1185

Nickerson DA, Tobe VO, Taylor SL (1997) Nucleic Acids Res **25:** 2745–2751

Oleykowski CA, Bronson Mullins CR, Godwin AK, Yeung AT (1998) Nucleic Acids Res **26:** 4597–4602

Pereira A, Aarts MG (1998) Methods Mol Biol **82:** 329–338

Que Q, Jorgensen RA (1998) Dev Genet **22:** 100–109

Redei GP, Koncz C (1992) *In* C Koncz, N-H Chua, J Schell, eds, Methods in Arabidopsis Research. World Scientific, Singapore, pp 16–82

Scherer S, Davis RW (1979) Proc Natl Acad Sci USA **76:** 4951–4955

Schimenti J, Bucan M (1998) Genome Res **8:** 698–710

Schuch W (1991) Symp Soc Exp Biol **45:** 117–127

Schy WE, Plewa MJ (1989) Mutat Res **211:** 231–241

Sega GA (1984) Mutat Res **134:** 113–142

Sharp PA (1999) Genes Dev **13:** 139–141

Smith D, Yanai Y, Liu YG, Ishiguro S, Okada K, Shibata D, Whittier RF, Federoff NV (1996) Plant J **10:** 721–732

Speulman E, Metz PL, van Arkel G, te Lintel Hekkert B, Stiekema WJ, Pereira A (1999) Plant Cell **11:** 1853–1866

Stadler LJ (1932) Proceedings of the VI Congress of Genetics **1:** 274–294

van Houwelingen A, Souer E, Spelt K, Kloos D, Mol J, Koes R (1998) Plant J **13:** 39–50

Waterhouse PM, Graham MW, Wang MB (1998) Proc Natl Acad Sci USA **95:** 13959–13964