High-Throughput Screening for Induced Point Mutations

Trenton Colbert, Bradley J. Till, Rachel Tompa, Steve Reynolds, Michael N. Steine, Anthony T. Yeung, Claire M. McCallum, Luca Comai, and Steven Henikoff*

Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109 (T.C., B.J.T., R.T., M.N.S., C.M.M., S.H.); Department of Botany, University of Washington, Seattle, Washington 98195 (S.R., L.C.); and Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111 (A.T.Y.)

With the completion of genome sequencing projects, emphasis in genomics has shifted from analyzing sequences to understanding gene function, and effective reverse genetic strategies are increasingly in demand. Here we report adaptations of the targeting induced local lesions in genomes (TILLING) reverse genetic strategy (McCallum et al., 2000a) to make it suitable for large-scale screening of chemically induced mutations in Arabidopsis and other plants.

TILLING has several advantages over other reverse genetic strategies. Unlike methods that provide only knockout mutations (e.g. Altmann et al., 1995), TILLING yields a traditional allelic series of point mutations. This will be especially valuable for essential genes, where sublethal alleles are required for phenotypic analysis. Because chemical mutagenesis causes a high density of mutations (Koornneef et al., 1982), virtually all genes can be targeted by screening relatively few individuals. Furthermore, the generality of chemical mutagenesis means that TILLING can be applied to plants without requiring transgenic or sophisticated tissue culture methodology. However, point mutations are relatively subtle changes, so their detection can be challenging. This problem has received much current attention because of the importance of discovering single nucleotide polymorphisms in humans for genotyping, and numerous strategies have been introduced (Kristensen et al., 2001). In the original TILLING method, we described the use of denaturing HPLC (dHPLC) for sensitive mutation discovery in pools (McCallum et al., 2000a). The availability of an automated column injector that accepts a 96-well microtiter plate (Underhill et al., 1997) meant that TILLING could be routinely performed in a central facility at the rate of about one gene per week. To obtain high throughput for genomic applications, we desired a reliable and inexpensive point mutation discovery method that could be performed more rapidly than dHPLC and in a robust manner.

GEL-BASED SCREENING FOR MISMATCHED HETERODUPLEXES

In the basic TILLING method (Fig. 1), seeds are mutagenized by treatment with EMS. The resulting M1 plants are self-fertilized, and the M2 generation of individuals is used to prepare DNA samples for mutational screening while their seeds are inventoried. DNA samples are pooled, and pools are arrayed on microtiter plates and subjected to gene-specific PCR. In the new high-throughput method described here, amplification products are incubated with an endonuclease that preferentially cleaves mismatches in heteroduplexes between wild type and mutant. Cleavage products are electrophoresed using an automated sequencing gel apparatus, and gel images are analyzed with the aid of a standard commercial image-processing program. Differential double end labeling of amplification products allows for rapid visual confirmation because mutations are detected on complementary strands, and therefore can be easily distinguished from amplification artifacts. Upon detection of a mutation in a pool, the individual DNA samples are similarly screened to identify the plant carrying the mutation. This rapid screening procedure determines the location of a mutation to within a few base pairs for PCR products up to 1 kb in size.

Several enzymes have been used for mismatch-specific cleavage, including S1 nuclease (Howard et al., 1999) and T4 endonuclease VII (Youil et al., 1996). We settled upon a recently described member of the S1 nuclease family, CEL I, a plant-specific extracellular glycoprotein (Oleykowski et al., 1998). CEL I has been shown to be suitable for genotyping applications because it preferentially cleaves mismatches of all types (Oleykowski et al., 1998) and has been used to detect heterozygous polymorphisms in DNA pools (Kulinski et al., 2000). Following PCR amplification of genomic DNA in 96-well plates, a solution containing CEL I is added and incubated. A stop solution is added and the mixture transferred to a 96-well Sephadex G50 spin plate for cleanup by centrifugation into a formamide-containing denaturation solution. After reducing the volume by heating, a robotic comb loader

[1] This work was supported by the National Science Foundation Plant Genome Program (grant to L.C. and S.H.), by the National Institutes of Health (grant nos. CA71426 to A.T.Y. and GM29009 to S.H.), and by the U.S. Department of Agriculture (grant no. 97–35301 to L.C.). S.H. is an Investigator of the Howard Hughes Medical Institute.

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

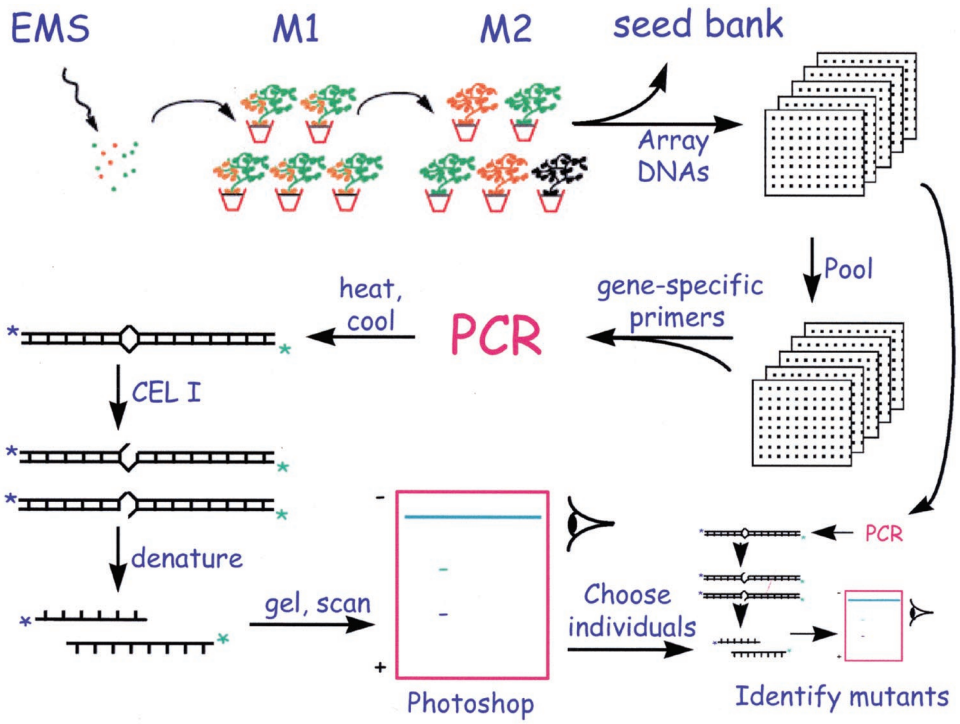

Figure 1. High-throughput TILLING. Starting with a single plant of Arabidopsis, ecotype Columbia homozygous for an *erecta* mutation (Torii et al., 1996), seeds were collected and mutagenized in batches at 20, 25, or 30 mM ethylmethanesulfonate (EMS) as described (McCallum et al., 2000a). M1 plants were allowed to grow in trays, and seeds were sown in pots for the M2 generation, where each M2 derived from a different M1 plant. M2 DNAs were prepared from 0.2 g of leaf and/or stem tissue using the Bio101 FastDNA system (http://www.qbiogene.com/protocols/dna-kits/p-fastdna.html) following the manufacturer's instructions, and concentrations were estimated by visualization on 1% (w/v) agarose electrophoretic gels and equalized before dilution (in 10 mM Tris [pH 8.0] and 1 mM EDTA) and (8-fold) pooling. PCR is performed in 10-μL volumes using ExTaq polymerase (Fisher/Panvera Labs), except that only half the manufacturer's recommended concentration of buffer is used, and $MgCl_2$ is increased to 2 mM. Primers are obtained from MWG Biotech (http://www.mwgbiotech.com/services/dna/index.htm) and mixed in a ratio of 3:2 (labeled:unlabeled) for the IR Dye 700-labeled primer and 4:1 (labeled:unlabeled) for the IR Dye 800-labeled primer, for final primer concentrations of 0.2 μM. Primers are designed with melting temperatures of 60°C to 70°C, and final annealing temperatures of melting temperature −5°C are chosen. Cycling is performed in MWG Biotech 96-well cyclers as follows: 95°C for 2 min; eight cycles of touchdown PCR (94°C for 20 s [denaturation], T_m + 3°C to T_m − 4°C decrementing 1°C per cycle [annealing], and 72°C for 45 s to 1 min [extension for 600–1,000-bp products]); 45 cycles of: 94°C for 20 s (denaturation), T_m − 5°C (annealing), and 72°C for 45 s to 1 min; 72°C for 5 min; 99°C for 10 min (inactivation); and 70 cycles of 20 s at 70°C to 49°C, decrementing 0.3°C per cycle (reannealing). Cycling is followed by CEL I treatment, cleanup, gel electrophoresis, and scanning (see Fig. 2).

transfers aliquots to a membrane comb, which is inserted into the well of a slab gel for electrophoresis.

Slab gel electrophoresis is well suited for large-scale mutation detection. The two-dimensional readout facilitates the detection of rare events, such as mutations, because a new band will stand out above the wild-type background and can be easily spotted. The size of each new band is also obtained, an advantage over other methods based on detection of mismatches or conformational changes (Nataraj et al., 1999), which do not indicate where in the molecule a mutation resides. So, although a new mutation in a coding exon will require sequencing, knowing its approximate location simplifies this step.

The DNAs are separated by denaturing gel electrophoresis and detected in two separate channels by scanners (LI-COR, Lincoln, NE; Middendorf et al., 1992). Sensitivity is sufficient to detect the approximately 100 attomoles of cleavage product generated by CEL I in an 8-fold pool, or one in 16 genomes for a heterozygous mutation. Opposed PCR primers carry different dye labels. Because there is no detectable overlap between the infrared (IR) Dye 700 and IR Dye 800 dye labels, images can be examined directly for the presence of novel bands in either channel. A UNIX perl program ("grab") retrieves and archives the image files from the LI-COR scanners via a file transfer protocol and processes them (using ImageMagick for UNIX, www.imagemagick.org) to create compressed JPEG files on a central server for Adobe Photoshop (Adobe, Seattle) analysis on networked local computers (Macintoshes and personal computers). A typical gel image (Fig. 2) will show a sequence-specific pattern of background bands resulting from endonucleolytic cleavages common to all 96 lanes. By superimposing images representing

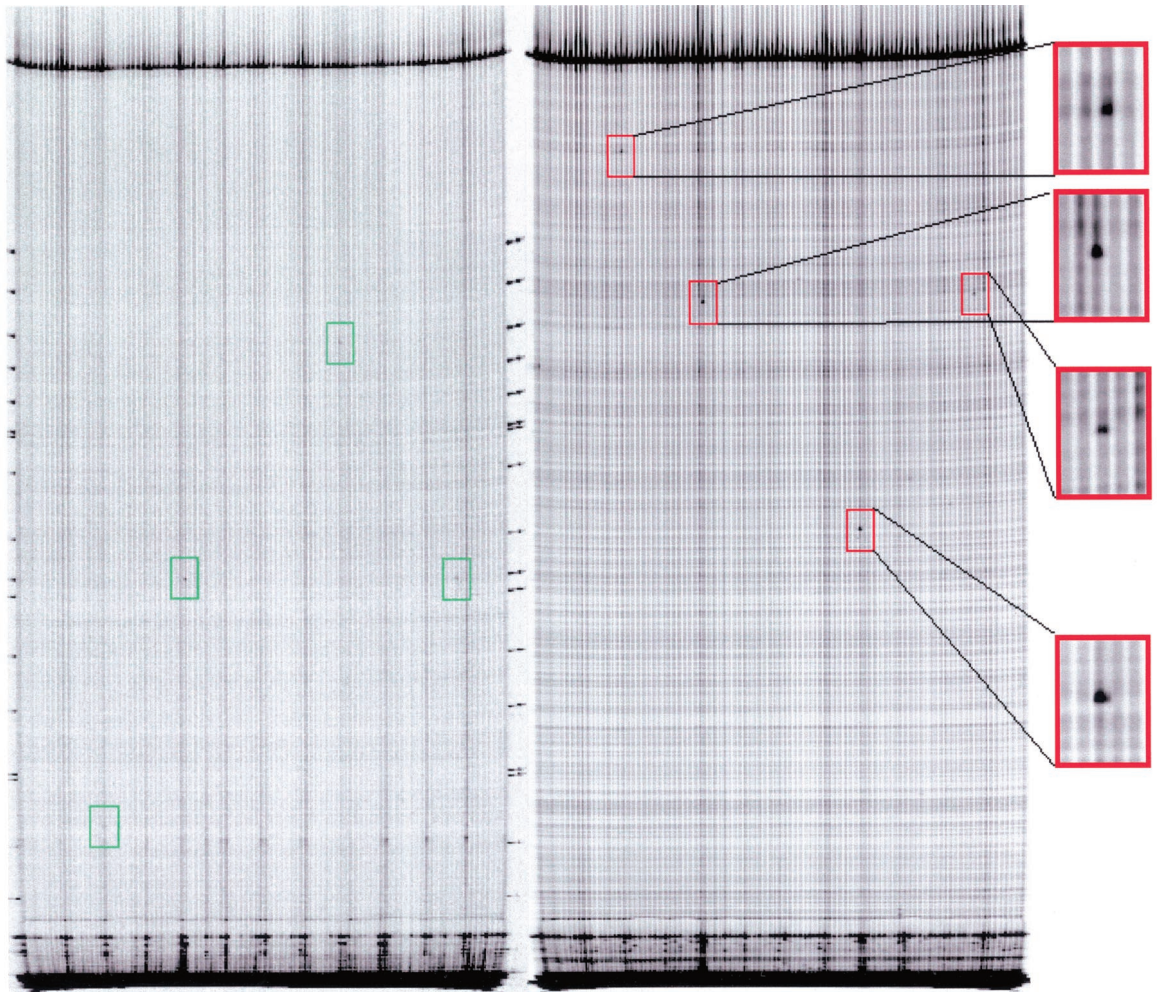

Figure 2. Example of a gel used for mutation detection in 8-fold pools. For digestion of 10-μL PCR products in 96-well plates, 20 μL of a solution containing 10 mM HEPES [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid] (pH 7.5), 10 mM $MgSO_4$, 0.002% (w/v) Triton X-100, 20 ng mL^{-1} of bovine serum albumin, and 1/1000 dilution of CEL I (50 units μL^{-1}) was added with mixing on ice, and the plate was incubated at 45°C for 15 min. CEL I was purified from 30 kg of celery as described by Oleykowski et al. (1998), except that Poros HQ rather than Mono Q was used, and the PhenylSepharose and Superdex 75 columns were omitted. The specific activity was 1×10^6 units mL^{-1}, where a unit is defined as the amount of CEL I required to digest 50% of 200 ng of a 500-bp DNA fragment that has a single mismatch in 50% of the duplexes. Reactions were stopped by addition of 5 μL 0.15 M EDTA (pH 8) and the mixture pipetted into wells of a spin plate (G50, Sephadex) prepared and spun according to the manufacturer's recommendations into a plate containing 1 to 1.5 μL of formamide load solution (1 mM EDTA [pH 8] and 200 μg mL^{-1} bromphenol blue in deionized formamide). The volume was reduced to a minimum by incubation at 96°C uncovered (30–40 min) and stored on ice, then transferred to a membrane comb using a comb loading robot (MWG Biotech). IR Dye 800-labeled M_r marker mix (50–700 bp) was applied to outside teeth. Following the prerun focusing step on a LI-COR Global IR^2 gel scanner, the comb was inserted, electrophoresed for 1 min, and removed. Electrophoresis was continued for 4 h at 1,500-V, 40-W, and 40-mA limits at 50°C. The figure shows IR Dye 800 (left) and IR Dye 700 (right) channels of a representative run. Bands corresponding to four of seven mutations detected on this gel are shown boxed, and sections of the IR Dye 700 images are magnified in offsets (far right). Note that these are seen only in one channel, but have counterparts in the other channel that add up to the length of the full-sized 1,012-bp product (band at top). Several bands near the bottom of the gel are detected in both channels: These artifactual bands result from random mispriming. A total of approximately 750 kb of sequence has been interrogated for point mutations on this single gel.

both channels and flipping between them, one can readily detect a lane containing a novel band in one channel and a corresponding novel band in the other channel. The sum of the two band sizes is equal to the full-length product visible at the top of the image. This visual assay is aided by the approximate proportionality of the migration distance to M_r, so that a band in one channel is nearly the same distance from the leading edge as the corresponding band in the other channel is from the full-length product. Photoshop image manipulation tools, rulers, and guides facilitate determination of migration distances and

lane numbers for the two bands. These data, together with subjective data quality assessments, are recorded using a perl program ("squint"), which applies a calibration curve to estimate M_rs.

An important advantage of double-end labeling for detecting both CEL I cleavage products is avoidance of false positive bands, of which there are two types: those that appear in multiple lanes in a single channel and those that appear in a single lane but in the same position in both channels. Because it is highly unlikely that the same mutation will appear in two different plants, we assume that certain homoduplex sites are especially sensitive to variability in CEL I digestion, causing bands to appear in multiple lanes above the background pattern. Bands that appear in both channels are likely to be examples in which mis-priming leads to a large amount of double-end-labeled product of a single size, with smaller products having a selective advantage over larger products during cycling, leading to sporadic low-M_r bands. We have found that PCR product yield is typically low and inconsistent using both IR Dye 700 and IR Dye 800 dyes on opposing primers; however, consistent results have been obtained using a mixture of IR Dye-labeled and unlabeled primers.

IDENTIFYING MUTATIONS IN POOLED AND INDIVIDUAL DNA SAMPLES

Initial experiments were performed using 5-fold pooling, which appears to be the practical limit of detection by dHPLC for fragments in the 500- to 600-bp range (McCallum et al., 2000a). By screening for mutations in the same fragments using both dHPLC and the current method, we could directly compare their detection levels. For example, we performed high-throughput TILLING on 5-fold pooled samples for the Sir2B gene, which had previously been carefully screened using dHPLC, with products confirmed by DNA sequencing. Six confirmed mutations, all heterozygous G/C to A/T transitions, were detected by both methods: Four were detected using dHPLC and five by the high-throughput method. When we increased pooling to 8-fold, we obtained similarly high detection levels without false positives: In one test, a screening of 4-fold pools found only the same seven mutations discovered in 8-fold pools of the same DNAs (data not shown). Therefore, we adopted an 8-fold pooling scheme.

Once a mutation is detected in a pool, the individual DNA samples comprising the pool are screened. Individual samples are arrayed in an 8×8 grid on microtiter plates, such that each pool corresponds to a row of individuals; thus, each column of the pool plate corresponds to a column of rows in the 8×8 grid. Using an eight-channel multipipettor, DNA is transferred from the row corresponding to the positive pool into a column of a fresh microtiter plate, so that 12 mutations per plate are screened as individuals. A UNIX perl program ("pick") facilitates this step by converting a set of squint output files into a table providing the plate number and row coordinate corresponding to each positive pool. To detect homozygotes as heteroduplexes, the individual samples are mixed with an equal amount of wild-type DNA. From this point on, screening to detect 12 individual mutations is identical to screening of pools, including amplification, CEL I digestion, gel electrophoresis, and grab, Photoshop, and squint analyses. This results in the identification of the plant in which a point mutation has occurred and an estimated location within a few base pairs of the lesion. Using this two-step strategy, we have been able to interrogate as much as approximately 750 kb of individual genomic sequences per gel (1 kb $\times$ eight plant DNAs $\times$ 96 lanes), and have identified mutations in 20 Arabidopsis chromatin genes (http://Ag.Arizona.Edu/chromatin/atgenes.html). For the most heavily mutagenized plants that we have screened, which displayed 30% embryo lethality after the first round of selfing, we estimate approximately seven point mutations per 8-fold pool plate (representing 768 plants) per gel for 1-kb fragments. This corresponds to approximately 1,000 EMS-induced mutations per Arabidopsis genome.

A HIGH-THROUGHPUT TILLING FACILITY

We are establishing a high-throughput facility for TILLING genes as a service to the Arabidopsis community and as a model for TILLING other plants. Using an interactive web-based system developed by Nick Taylor and Elizabeth A. Greene (Fred Hutchinton Cancer Research Center, Seattle; http://www.proweb.org/coddle), a user chooses primers designed to maximize the probability of a deleterious mutation in the gene of interest. Regions that have high stop codon potential and high evolutionary conservation are most useful for providing an allelic series (McCallum et al., 2000b). By our current plan, users will be billed directly by an oligonucleotide manufacturer for primers, which will be delivered to the facility for screening pools and individuals at no charge. Upon discovery of mutations in reference plants, the user will report the plant number and the position of the mutation to the facility electronically. The user will also receive seeds from the mutant plants via the Arabidopsis Biological Resource Center (Columbus, OH) for a nominal charge. We also anticipate providing aliquots of primers and DNAs from mutant plants sufficient for amplification and base determination either by DNA sequencing or by direct termination PCR, which has recently been adapted for the LI-COR double end-labeling system (Chen et al., 2001). In return, the user will be expected to ascertain the base change in each mutant, which is usually a G/C to A/T transition (McCallum et al., 2000b), and report it back to the TILLING project,

where it will be entered in a public mutant database, accessible by BLASTN and BLASTP analysis. Because our gel-based detection system determines the location of the mutation to within a few base pairs, the task of identifying the precise mutation is greatly simplified, especially for heterozygotes, where mutant and wild-type sequences overlap. Depending on whether the mutant plant is homozygous or heterozygous, the user will analyze one or more DNAs by amplification and sequencing or typing, and perform the phenotypic analyses and subsequent crosses (McCallum et al., 2000b). It is expected that most mutations will be missense alleles. For mutations that fall into conserved regions of proteins, it is possible to predict their severity using the recently introduced SIFT algorithm (Ng and Henikoff, 2001), which is available for interactive use on the web (http://blocks.fhcrc.org/~pauline/SIFT.html).

Our high-throughput procedure is rapid and relatively inexpensive. A single technician can easily perform all operations at the rate of four gel runs per day, enough to screen for mutations in 3,000 plants. For the highest mutation rates that we have obtained (approximately 1,000 per genome), this corresponds to more than 20 mutations, enough to provide a better than even chance of at least one knockout lesion in a typical gene, plus an allelic series of a dozen or more missense mutations. With standard 96-well pipettors and robotics replacing manual multipipettors and PCR machines, we expect that the capacity of our four LI-COR scanners can be increased to 16 runs per day, enough to TILL at the rate of three to four genes per day.

CONCLUSIONS

By taking advantage of robust equipment developed for high-throughput sequencing and genotyping and a popular image analysis program developed for the general public, we have been able to streamline plant reverse genetics. We expect that TILLING will be comparably efficient for plants with larger genomes because EMS toxicity is expected to scale with the number of functional genes, which is likely to be similar for all higher plants. The generality of our methodology encourages its application beyond plants. For example, our high-throughput methodology should be directly applicable to an EMS-based reverse genetic method utilizing dHPLC described for *Drosophila melanogaster* (Bentley et al., 2000). Because there is no practical method for maintaining fertile *D. melanogaster* beyond several weeks, a rapid procedure is especially desirable to minimize the burden of continually mutagenizing and culturing flies. Another possible application of high-throughput TILLING is in an ongoing *Caenorhabditis elegans* reverse genetics project that provides knockouts using chemical mutagenesis for production of deletions (Jansen et al., 1997). Given the utility of allelic series to complement knockout and RNA-mediated inhibition analyses, we can envision the adoption of our procedures for worms as well. Thus, there is the prospect that a technology introduced in plants will be adopted for these preeminent model organisms, a reversal of recent trends in genomics.

ACKNOWLEDGMENTS

We thank Jorja Henikoff for writing grab, squint, and pick; Elizabeth Greene and Nick Taylor (FHCRC) for making CODDLE available to us prior to publication; Jochen Jaeger and Ann Slade for helpful discussions; and Michelle Acupanda, Brianna Borders, Amy Holmes, Jessica Johnson, Christine Codomo, Amber Kost, and Kim Young for planting, harvesting, and preparing the DNA samples.

Received March 6, 2001; returned for revision March 13, 2001; accepted March 16, 2001.

LITERATURE CITED

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

Bentley A, MacLennan B, Calvo J, Dearolf CR (2000) Genetics **156:** 1169–1173

Chen JZ, Smith L, Pfeifer GP, Holmquist GP (2001) Nucleic Acids Res **29:** e17–e21

Howard JT, Ward J, Watson JN, Roux KH (1999) Biotechniques **27:** 18–19

Jansen G, Hazendonk E, Thijssen KL, Plasterk RH (1997) Nature Gen **17:** 119–121

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

Kristensen VN, Kelefiotis D, Kristensen T, Borresen-Dale A (2001) Biotechniques **30:** 318–332

Kulinski J, Besack D, Oleykowski CA, Godwin AK, Yeung AT (2000) Biotechniques **29:** 44–46

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

McCallum CM, Comai L, Greene EA, Henikoff S (2000b) Plant Physiol **123:** 439–442

Middendorf LR, Bruce JC, Bruce RC, Eckles RD, Grone DL, Roemer SC, Sloniker GD, Steffens DL, Sutter SL, Brumbaugh JA (1992) Electrophoresis **13:** 487-494

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

Ng PC, Henikoff S (2001) Genome Res (in press)

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

Torii KU, Mitsukawa N, Oosumi T, Matsuura Y, Yokoyama R, Whittier RF, Komeda Y (1996) Plant Cell **8:** 735–746

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Genome Res **7:** 996–1005

Youil R, Kemper B, Cotton RG (1996) Genomics **32:** 431–435