Published in final edited form as:

Methods Enzymol. 2005 ; 395: 570–596.

Methods for Genome-Wide Analysis of Gene Expression Changes in Polyploids

Jianlin Wang, Jinsuk J. Lee, Lu Tian, Hyeon-Se Lee, Meng Chen, Sheetal Rao, Edward N. Wei, R. W. Doerge, Luca Comai, and Z. Jeffrey Chen

Abstract

Polyploidy is an evolutionary innovation, providing extra sets of genetic material for phenotypic variation and adaptation. It is predicted that changes of gene expression by genetic and epigenetic mechanisms are responsible for novel variation in nascent and established polyploids (Liu and Wendel, 2002; Osborn *et al.*, 2003; Pikaard, 2001). Studying gene expression changes in allopolyploids is more complicated than in autopolyploids, because allopolyploids contain more than two sets of genomes originating from divergent, but related, species. Here we describe two methods that are applicable to the genome-wide analysis of gene expression differences resulting from genome duplication in autopolyploids or interactions between homoeologous genomes in allopolyploids. First, we describe an amplified fragment length polymorphism (AFLP)–complementary DNA (cDNA) display method that allows the discrimination of homoeologous loci based on restriction polymorphisms between the progenitors. Second, we describe microarray analyses that can be used to compare gene expression differences between the allopolyploids and respective progenitors using appropriate experimental design and statistical analysis. We demonstrate the utility of these two complementary methods and discuss the *pros* and *cons* of using the methods to analyze gene expression changes in autopolyploids and allopolyploids. Furthermore, we describe these methods in general terms to be of wider applicability for comparative gene expression in a variety of evolutionary, genetic, biological, and physiological contexts.

AFLP-cDNA Display

Introduction

Amplified fragment length polymorphism (AFLP) (Vos *et al.*, 1995) is used to develop DNA markers for genome mapping and genetic diversity analysis. The method can be used to analyze simultaneously different DNA regions throughout the entire genome regardless of its origin or complexity (Mei *et al.*, 2004). A modification of the method using complementary DNA (cDNA) samples in the analysis, known as AFLP-cDNA or cDNA-AFLP display, allows the characterization of tissue-specific gene expression patterns (Bachem *et al.*, 1996) and detection of gene expression differences in allopolyploids (Comai *et al.*, 2000; Lee and Chen, 2001; Madlung *et al.*, 2002). The technique employs the use of restriction polymorphism in analyzing cDNA samples (Fig. 1A). After the cDNA fragments are digested, they are ligated onto adapters with compatible restriction ends. Ligated products are subjected to polymerase chain reaction (PCR) analysis using two steps of preselective and selective amplification so that subsets of transcripts are amplified and resolved in a sequencing gel (Sambrook *et al.*, 1989). Any fragments that are present in one sample but not the other are isolated and sequenced (Lee and Chen, 2001). The method is particularly useful to detect expression changes in orthologous and homoeologous loci that display polymorphism in the restriction sites of the enzymes used in the assay. Compared with the traditional messenger RNA (mRNA) display (Liang and Pardee, 1992) and random arbitrarily primed PCR (RAP-PCR) (Welsh *et al.*, 1992), the AFLP-cDNA display is relatively sensitive and reproducible. The technique is sensitive enough to detect weakly expressed transcripts, such as genes encoding transcription factors that are undetectable by Northern blot analysis (Lee and Chen, 2001). Moreover, it does not detect as many false positives as observed in regular mRNA display. Finally, it can be used

for quantitative detection because the intensity of each fragment reflects the expression level of the gene relative to internal controls (e.g., constantly expressed cDNAs), allowing measurement of dosage-dependent gene regulation in autopolyploid series.

Another advantage of using AFLP-cDNA display is that there is no prerequisite for the DNA sequence information. Novel transcripts are readily detectable. The origin of the expressed homoeologous transcripts may be determined by comparing the polymorphic cDNA fragments between the allotetraploids and their progenitors. Moreover, the method can be applied to any organisms, especially those whose genomic resources are underdeveloped. By comparing the sequences of the cloned fragments with known sequences in the databases, one might identify homologous genes in the related species. However, the efficiency of AFLP-cDNA display is limited by *post hoc* cloning and sequencing of individual cDNA fragments. False positives may be detected because of PCR artifacts or heterogeneous populations (e.g., outcrossing plants). The comparison of gene expression patterns (absence or presence of a fragment) between an allopolyploid and its progenitors may fail to distinguish gene expression differences because of allelic variation among individual plants within a progenitor species. Therefore, to reduce the number of false positives, it is recommended to use biological replications and to verify the gene expression patterns detected in AFLP-cDNA display by other methods such as reverse-transcriptase PCR (RT-PCR) and cleaved amplified polymorphic sequence (CAPS), quantitative RT-PCR, sequencing, and/or SSCP (Adams *et al.*, 2003; Lee and Chen, 2001; Lee *et al.*, 2004), and other methods as described in Chapter 29.

Protocol

The protocol for AFLP-cDNA display is modified from the published methods (Bachem *et al.*, 1996; Comai *et al.*, 2000; Lee and Chen, 2001) and recommendations from the manufacturer (Invitrogen, Carlsbad, CA). The protocol includes six steps (Fig. 1A). They are (1) isolate mRNA; (2) synthesize double-stranded (ds) cDNA; (3) generate cDNA fragments by restriction digestion; (4) ligate the cDNA fragments onto adapters that contain the restriction enzyme sites; (5) amplify subsets of cDNA fragments using two-step (preselective and selective) amplification; and (6) display the cDNA fragments in a sequencing gel, clone, and sequence the cDNA fragments that are differentially expressed between the two samples. In this experiment (Fig. 1B), we prepared RNA samples from an *Arabidopsis thaliana* diploid (At), an autotetraploid (At4), *Arabidopsis arenosa* autotetraploid, four selfing generations (S2–S5) of a new allotetraploid (745), and natural allotetraploid *Arabidopsis suecica* (As), to demonstrate the detection of changes of gene expression in diploid and isogenic autotetraploid and allotetraploid lines. In addition to the results shown in Fig. 1B, detailed analyses of gene expression patterns in these lines are described elsewhere (Wang *et al.*, 2004).

Isolation of Total RNA

All reagents are purchased from Invitrogen, unless noted otherwise. Every solution is made by diethylpyrocarbonate (DEPC)-treated double-deionized autoclaved water (ddH₂O) (Sambrook *et al.*, 1989). For each line, total RNA is extracted from fresh leaves or flower buds of pools of 10 – 20 plants (3 – 4 weeks old) using the TRIzol reagent (Invitrogen). In a typical preparation, about 500 µg of total RNA can be obtained from 1 g of fresh leaves or flower buds in *Arabidopsis* and *Brassica*.

Grind the plant samples (1 g) in liquid nitrogen and transfer the fine powders to a 30-ml centrifuge tube. Add 10 ml of TRIzol reagent, mix well by vortexing for 20 s, and incubate in a waterbath at 65° for 5 min. After cooling, add 2 ml of chloroform and mix vigorously for 20 s. Incubate at room temperature (RT) for 5 min followed by centrifugation at 10,000g for 15 min at 4°. Transfer the supernatant to a 50-ml polypropylene conical tube, add about 1 volume of isopropyl alcohol, and incubate at RT for 45 min. Centrifuge at 3000 rpm for 15 min at 4°

to precipitate RNA. Wash the RNA pellet twice using 70% ethanol, briefly dry the RNA pellet, and dissolve the pellet in 300 μ l of DEPC-treated ddH₂O or RNase-free TE buffer (10 mM Tris-HCl and 1 mM ethylenediaminetetraacetic acid [EDTA], pH 8.0).

Isolation of mRNA

mRNA is isolated using the Micro-FastTrack 2.0 mRNA isolation kit (Invitrogen). Approximately 5 μ g of poly(A) RNA can be purified from about 500 μ g of total RNA using 25 mg of oligo-(dT) cellulose.

Add 1 ml of binding buffer (500 mM NaCl, 10 mM Tris-Cl, pH 7.5) to the total RNA solution, heat it at 65° for 5 min, and place it immediately on ice for exactly 1 min.

Transfer the solution to a 2-ml microcentrifuge tube containing 25 mg of oligo-(dT) cellulose and incubate at RT for 2 min. Then gently rotate the tube for 0.5–1.0 h at RT to increase the efficiency of mRNA binding to oligo-(dT) cellulose. Centrifuge at 4000g for 5 min at RT and carefully remove the supernatant (do not disturb the cellulose pellet).

Resuspend the oligo-(dT) cellulose in 1.3 ml of binding buffer and centrifuge at 4000g for 5 min and carefully remove the supernatant. This step is repeated (at least three times) until the buffer is no longer cloudy.

Resuspend the oligo-(dT) cellulose in 0.3 ml of binding buffer and transfer it to a spin column. Centrifuge at 4000g for 10 s at RT and carefully remove the supernatant. Repeat this washing procedure using 0.5 ml of binding buffer for at least three times until the OD₂₆₀ of the flow through is 0.05 or less.

To the oligo-(dT) cellulose, add 200 μ l of low-salt washing buffer (250 mM NaCl, 10 mM Tris-Cl, pH 7.5) and gently mix. Centrifuge the column and remove the liquid as described earlier. Repeat washing with additional 200 μ l of low-salt washing buffer for at least three times. Washing using low-salt buffer helps remove sodium dodecyl sulfate (SDS) and non-polyadenylated RNAs such as rRNAs.

Place the spin column into a new microcentrifuge tube, add 100 μ l of elution buffer, and mix gently by pipetting. Centrifuge 10 s at RT. Repeat the procedure using another 100 μ l of elution buffer. The final volume of mRNA in the elution buffer is 200 μ l.

Quantify mRNA in a 1:1 dilution (5 μ l of mRNA sample and 5 μ l of elution buffer) using a spectrophotometer (GeneQuant *pro*, Amersham Bioscience, Piscataway, NJ). Use 10 μ l of elution buffer to calibrate the spectrophotometer.

To the about 200 μ l of mRNA solution, add 10 μ l of 2 mg/ml of glycogen, 30 μ l of 3 M Na-acetate, 600 μ l of 100% ethanol, and mix. Store at –70° until use.

Place the samples at RT for 15 min. Centrifuge at 14,000 rpm for 20 min at 4°. Remove the ethanol and wash the pellet with 80% ethanol. Centrifuge briefly, remove ethanol, and dissolve the mRNA pellet in the elution buffer to a final concentration of 100 ng/ μ l. The mRNA can be used immediately or stored at –70°.

First-Strand cDNA Synthesis

The cDNAs are synthesized using a slightly modified procedure from that provided by the universal RiboClone cDNA synthesis system (Promega, Madison, WI). To synthesize the first-strand cDNA, add 2 μ l oligo-(dT)_{12–18} (500 μ g/ml) and 0.5–1.0 μ g of mRNA to a 0.5-ml RNase-free microcentrifuge tube. Adjust the volume to 15 μ l using nuclease-free ddH₂O. Heat the

reaction mixture at 70° in a thermal cycler for 10 min and tube on ice. After brief centrifugation, add 5 µl of first-strand 5× buffer, 1 µl of RNasin ribonuclease inhibitor (40 U/µl), 2.5 µl of sodium pyrophosphate (40 mM), and 2.5 µl of AMV reverse transcriptase (24 U/µl) to each tube. Mix the solution by gently pipetting up and down several times and incubate at 42° for 1.5 h.

Second-Strand cDNA Synthesis

The procedure is performed immediately after the previous reaction. Add 40 µl of second-strand 2.5× buffer, 2.5 µl of DNA polymerase I (9 U/µl), 0.8 µl of RNase H (2 U/µl), 5 µl of acetylated bovine serum albumin (BSA) (1 mg/ml), and 26.7 µl of nuclease-free ddH₂O. Gently mix the solution in a total volume of 100 µl and incubate at 14° for 2 h. Heat this reaction to 70° for 10 min and place it on ice. Add 1 µl of T4 DNA polymerase (7.7 U/µl) and incubate it at 37° for 10 min. The reaction is terminated by the addition of 4 µl of 200 mM EDTA. Place the tube on ice or store at −20°.

The size of cDNA should be examined after the second-stranded cDNA synthesis. Take 5 µl of the reaction solution (and 0.5 µl 10× loading buffer) and load it onto a 1% agarose gel. After electrophoresis, the DNA can be stained using SYBR gold dye (Molecular Probes, Inc., Eugene, OR) and visualized under ultraviolet (UV) light. A successful reaction generates a smear of DNA with a range of 500 – 4000 bp in length.

The cDNAs are purified by phenol–chloroform extraction and ethanol precipitation (Sambrook *et al.*, 1989). To the sample tube, add an equal volume (~100 µl) of phenol:chloroform:isoamyl alcohol (25:24:1) saturated with TE buffer (pH 8.0) and mix it by vortexing. Centrifuge the tubes at 14,000 rpm in a microcentrifuge tube for 2 min and transfer the aqueous phase of the sample to a new tube. Add 0.1 volume of 2.5 M sodium acetate (pH 5.2) and 2 volumes of chilled (−20°) 100% ethanol, and then place the tube at −20° for 30 min. After centrifugation at 14,000 rpm for 20 min, the supernatant of the sample is carefully removed and the pellet is rinsed with 0.5 ml of cold (−20°) 70% ethanol. After a brief centrifugation, the 70% ethanol is carefully removed and the pellet is dried at RT. The cDNA pellet is dissolved in 10–50 µl of TE buffer and ready for restriction digestion or stored at −20° for future use.

Restriction Digestion

The double-stranded cDNAs are digested using two restriction endonucleases together to generate DNA fragments for ligation and amplification. In general, two restriction enzymes with 6- and 4-bp recognition sites, respectively (e.g., *EcoR* I and *Mse* I) are used to digest cDNA so that a relatively large number of polymorphic DNA fragments can be generated. Customized primers for adapter ligation and PCR amplification can be synthesized if other restriction enzymes (e.g., *Hind* III instead of *EcoR* I) are used. The combination of these two restriction endonucleases generates an optimum spectrum of fragments (<1 kb) for amplification and separation on a polyacrylamide gel. The completion of restriction enzyme digestion is very important as partial digestion of the cDNA generates ambiguous polymorphism that is not caused by the differences in cDNA or transcript profiles.

To digest cDNA, add 5 µl of 5× reaction buffer, about 18 µl of cDNA (~250 ng), 2 µl of *EcoR* I/*Mse* I (1.25 U/µl each) as supplied by the manufacturer (Invitrogen) to a final volume of 25 µl. Mix gently, centrifuge briefly, and incubate the reaction at 37° for at least 2 h. The reaction is inactivated by heating at 70° for 15 min.

Ligation of Adapters

The digested cDNA fragments are ligated onto double-stranded oligo DNA fragments that contain *EcoR* I and *Mse* I recognition sites (Fig. 1A). These short DNA fragments are called

EcoR I and *Mse* I adapters. The cDNA fragment ends with specific sequences can be used to design primers for PCR amplification, which prevents the nonspecific amplifications observed in regular mRNA displays (Liang and Pardee, 1992). Thus, a high efficiency of ligation is as important as the completion of restriction enzyme digestion so that the amplified fragments represent the expression profile differences among the biological samples tested.

To the reaction containing digested cDNAs, add 24 μ l of adapter ligation solution (*EcoR* I adapter, 5 μ M and *Mse* I adapter, 50 μ M), 0.4 mM ATP, 10 mM Tris-HCl, pH 7.5, 10 mM magnesium acetate, and 50 mM potassium acetate) and 1 μ l of T4 DNA ligase (1 U/ μ l, Invitrogen). Centrifuge briefly and incubate at 20° for 2 or more hours. After ligation, the reaction mixture can be diluted to 5, 10, or 15 times for PCR amplification. Store the unused portion of the reaction at -20°.

PCR Amplification

PCR is performed using two consecutive steps, namely, preselective and selective amplification. Ligated cDNA template is used for preselective amplification, in which the *Mse* I primer has one selective nucleotide at the 3' end, whereas the *EcoR* I primer contains no selective nucleotide, so only a fraction of ligated cDNA fragments are amplified. The PCR products from preselective amplification are diluted and used for selective amplification using another primer pair, an *Mse* I primer containing three selective nucleotides and an *EcoR* I primer containing two selective nucleotides. The combination of primer pairs used in each amplification should be tested empirically or according to the recommendations of the manufacturer (Gibco BRL/Invitrogen). The two-step amplification strategy results in a high-resolution sequencing gel with reproducible cDNA fragments produced from thousands of AFLP reactions. Moreover, the primer design and selective amplification strategy preferentially amplify *Mse* I-*EcoR* I fragments rather than *Mse* I-*Mse* I and *EcoR* I-*EcoR* I fragments.

Preselective Amplification

To a 0.2-ml thin-walled microcentrifuge tube, add 4 μ l of diluted template DNA (from the ligation reaction), 40 μ l of preselective primer mix (supplied by the manufacturer), 5 μ l of 10 $\times$ PCR buffer (including 15 mM MgCl₂), and 1 μ l of *Taq* DNA polymerase (5 U/ μ l) in a final volume of 50 μ l. Mix and centrifuge briefly. PCR is performed for 20 cycles of 94° for 30 s, 56° for 60 s, and 72° for 60 s, with a final extension at 72° for 10 min. The PCR products are diluted by 1:25 (transfer 5 μ l of the PCR products into a new 1.5-ml tube containing 120 μ l of ddH₂O).

An optional step monitors the results of preselective amplification. Label the *EcoR* I primer (without selective nucleotide) with [γ -³³P]ATP (see later discussion) and add 5 μ l of the labeled primer and 35 μ l preselective primer mix to the PCR. After PCR amplification, an aliquot of the PCR products can be electrophoresed in a polyacrylamide gel. The gel is dried in a vacuum gel dryer (Bio-Rad, Hercules, CA) at 80° for 2 h and exposed to an X-ray film or a PhosphorImager (Fujifilm Medical Systems, Inc., Stamford, CT). Successful amplification from the ligated cDNA templates generates a smear of DNA fragments ranging from 50 to 700 bp. No visible smearing above 2 kb or around the well should be observed. However, the products of abundant transcripts may be visible as prominent bands. Therefore, it is recommended to dilute (1:20–50) the amplified products to about 1 ng/ μ l for selective amplification.

Selective AFLP Amplification

Selective amplification is performed using primer pairs that contain two to three selective nucleotides in the 3' end of each primer. The introduction of selective bases reduces the number

of *EcoR* I–*Mse* I fragments that can be amplified during PCR. The number of all possible primer combinations is 1024 (16×64) if two selective bases are used in one primer and three in another. However, not every primer pair is useful for PCR amplification. Among all primer pairs tested by the manufacturer (Gibco BRL), 64 pairs are recommended for selective amplification in small genome-size species like *Arabidopsis* and *Brassica*. These include (1) eight *EcoR* I primers, each with two selective nucleotides in the 3' end: 5'-GAC TGC GTA CCA AAT CAA-3' (E-AA), CAC-3' (E-AC), CAG-3' (E-AG), CAT-3' (E-AT), CTA-3' (E-TA), CTC-3' (E-TC), CTG-3' (E-TG), and CTT-3' (E-TT); and (2) eight *Mse* I primers, each with three selective nucleotides: 5'-GAT GAG TCC TGA GTA ACA A-3' (M-CAA), ACAC-3' (M-CAC), ACAG-3' (M-CAG), ACAT-3' (M-CAT), ACTA-3' (M-CTA), ACTC-3' (M-CTC), ACTG-3' (M-CTG), and ACTT-3' (M-CTT). For large-genome species such as cotton and maize, a different set of *EcoR* I primers is recommended. The eight *EcoR* I primers are as follows: 5'-GAC TGC GTA CCA AAT CAA C-3' (E-AC), AAG-3' (E-AG), ACA-3' (E-CA), ACT-3' (E-CT), ACC-3' (E-CC), ACG-3' (E-CG), AG C-3' (E-GC), and AGG-3' (E-GG). The *Mse* I primers are the same for both small- and large-genome species.

Primer Labeling

In a 1.5 microcentrifuge tube, add 9 μ l of an *EcoR* I primer (27.8 ng/ μ l), 5 μ l of ddH₂O, 5 μ l of 5 $\times$ kinase buffer, 1 μ l of T4 polynucleotide kinase (10 U/ μ l, Invitrogen), and 5 μ l of [γ -³³P] ATP to a final volume of 25 μ l. Mix and centrifuge briefly and incubate the reaction at 37° for 1 h. The reaction is inactivated by heating at 70° for 10 min.

For “mix 1,” add 2.5 μ l of the [γ -³³P] ATP-labeled *EcoR* I–primer and 22.5 μ l of the *Mse* I–primer mix (6.7 ng/ μ l) that contains deoxyribonucleic triphosphates (dNTPs) to the final volume of 25 μ l.

For “mix 2,” add 39 μ l of ddH₂O, 10 μ l of 10 $\times$ PCR buffer (including 15 mM MgCl₂), and 1 μ l of *Taq* DNA polymerase (5 U/ μ l) to the final volume of 50 μ l.

Add 5 μ l of mix 1 (primers, dNTPs) and 10 μ l of mix 2 to a 0.2-ml thin-walled microcentrifuge tube containing 5 μ l of diluted PCR products from preselective amplification. Mix gently and centrifuge briefly before the tubes are placed in a thermal cycler. PCR is performed for one cycle of 94° for 30 s, 65° for 60 s, and 72° for 60 s plus 12 cycles of amplification with a descending annealing temperature of -0.7° in each cycle, which is followed by additional 23 cycles of 94° for 30 s, 56° for 60 s, and 72° for 60 s, with a final extension at 72° for 10 min. A high annealing temperature (at 65°) in the first PCR cycle and a gradient decrease in the annealing temperature during the subsequent 12 cycles provide a high level of selectivity for the primers containing selective nucleotides during PCR amplification.

AFLP-cDNA Analysis in a Polyacrylamide Gel

Many sequencing gel systems are appropriate for AFLP-cDNA display. We use a Gibco BRL (Invitrogen) or Bio-Rad sequencing apparatus, in which 49 or 97 samples can be loaded using 0.4-mm spacers and shark-tooth combs.

Make a 5% polyacrylamide gel solution containing 5% acrylamide:bisacrylamide (19:1), 7.5 M urea, and 1 $\times$ TBE buffer (89 mM Tris, 89 mM Borate, 2 mM EDTA, pH 8.0) (Sambrook *et al.*, 1989). Add 200 μ l of fresh 10% ammonium persulfate (APS) and 50 μ l of N,N,N',N'-tetramethyl-ethylenediamine (TEMED) to 80 ml of gel solution for polymerization. Pour the gel solution quickly into a sequencing gel unit.

Add 20 μ l of formamide loading dye (98% formamide, 10 mM EDTA, bromophenol blue, xylene cyanol) to 20 μ l of the PCR-amplified products. Heat at 90° for 3 min and chill the tube on ice before loading.

Pre-electrophoresis the gel until the gel temperature is about 45° (45 Watts for 30 min to 1 h). Load 5 μ l of each sample to the gel. Electrophoresis the gel at 80 Watts (constant wattage) until the xylene cyanol dye is 1/2–2/3 down the length of the gel (when the bromophenol blue runs off the gel), which usually results in a display of fragment sizes from 50 to 1000 bp.

After electrophoresis, transfer the gel onto a Whatman 3MM paper and dry the gel in a vacuum gel dryer (Bio-Rad) at 80° for 2 h. Expose the gel to a Kodak X-ray film at –80° overnight and develop the film in an automatic film developer (SRX-101A, Konica) or expose the gel to a screen for 3–8 h and detect the images in a PhosphorImager (Fujifilm Medical Systems, Inc.). The AFLP-cDNA profiles are clearly visible in the autoradiogram (Fig. 1B) or in a TIFF file (data not shown).

Reamplification of the Differentially Expressed cDNA Fragments

To recover the DNA fragments from the gel, completely dry the DNA sequencing gel and clearly mark each corner of the gel using a permanent marker. Alternatively, staple the gel to the film or poke a few “guide holes” using a needle.

After the film is developed, the autoradiogram is “perfectly” aligned with the gel using the marks in each corner, the staple, or guide holes. Use sturdy (bulldog) clips to hold the gel and film and place the film directly towards a light box. The bands should be visible through the gel and Whatman 3M paper. Use a sharp razor blade to slice out the bands of interest. Clearly mark the band locations on the autoradiogram.

Place the gel slice in 10 μ l of TE on a piece of parafilm for 3–5 min and carefully remove the 3M paper attached to the gel using a pair of fine tweezers. Transfer the gel slice and TE into a 1.5-ml microcentrifuge tube containing 90 μ l of TE. Pierce a hole in the lid and boil the tube for 15 min. Centrifuge at 14,000 rpm for 10 min and transfer the supernatant to a new tube. Add 10 μ l of 3M sodium acetate, 5 μ l of glycogen (10 mg/ml), and 300 μ l of 95% ethanol to the solution and incubate at –20° for 2 h. Centrifuge at 14,000 rpm for 20 min at 4°, remove the supernatant, and wash the DNA pellet using 500 μ l of cold 70% ethanol. The DNA pellet is dried at RT and dissolved in 20 μ l of ddH₂O. An aliquot of 5 μ l is used for reamplification of the DNA fragment using PCR. The protocol for cloning and sequencing the amplified PCR products is described elsewhere (Lee and Chen, 2001). After the sequence and its coding genes are determined, the expression levels of the gene are verified using RT-PCR, CAPS, and/or SSCP analysis (Adams *et al.*, 2003; Lee and Chen, 2001; Lee *et al.*, 2004).

Notes

1. It is extremely important to collect the samples or tissues from plants growing in the same conditions for gene expression analysis because expression profiles may change in variable growth conditions. Biological replications (RNAs isolated from two different pools of plants of the same genotype or treatment) should be used whenever possible. Protocols for germinating seeds and growing plants are described elsewhere (Comai *et al.*, 2000; Lee and Chen, 2001; Lee *et al.*, 2004; Madlung *et al.*, 2002).
2. The completion of restriction digestion is critical for AFLP analysis, which can be monitored using controls in which genomic and plasmid DNAs are digested using the same enzymes.
3. Adapter ligation is another critical process that should be monitored using an aliquot of ligation mix for PCR amplification. A successful ligation will result in a clear smear of DNA fragments between 50 and 600 bp.
4. Primer pairs used for preselective and selective amplification may be empirically tested. Alternatively, *Eco*R I and its primers may be replaced with another restriction

enzyme such as *Hind* III or *Dra* I. *EcoR* I activity is sensitive to DNA methylation, which may affect the genomic AFLP results. However, data obtained from our lab and others indicate that similar polymorphism levels of the fingerprints may be obtained using restriction enzymes other than *EcoR* I (Mei *et al.*, 2004).

5. Glassware, disposable containers, and solutions for RNA samples should be RNase free as recommended (Sambrook *et al.*, 1989). Some plasticware that cannot be autoclaved may be treated using special detergents (e.g., RNase Away, Molecular Bioproducts, San Diego, CA).
6. AFLP analysis can be automated using a LI-COR sequencer (LI-COR Biosciences, Lincoln, NE) if the primers are labeled using IRDyes or other fluorescent dyes (Klein *et al.*, 2000) instead of [γ -³³P]ATP.

Spotted 70-mer Oligonucleotide (Oligo) Microarrays

Introduction

Although AFLP-cDNA display has the advantage of discriminating expressed homoeologous transcripts between the allotetraploids and their progenitors (Comai *et al.*, 2000; Lee and Chen, 2001; Madlung *et al.*, 2002; Wang *et al.*, 2004), the efficiency is limited by *post hoc* cloning and sequencing of individual cDNA fragments. Other methods such as serial analysis of gene expression (SAGE) (Velculescu *et al.*, 1995) and massively parallel signature sequencing (MPSS) (Brenner *et al.*, 2000) also can be used for genome-wide analysis of transcription profiles. However, the techniques require a series of molecular manipulations.

DNA microarrays (Fodor *et al.*, 1991; Hughes *et al.*, 2001; Lipshutz *et al.*, 1999; Schena *et al.*, 1998; Singh-Gasson *et al.*, 1999) are applicable to transcriptome analysis in polyploid series. They offer advantages over AFLP-cDNA display and other methods because all gene sequences can be spotted simultaneously and the sequences of the spots on the microarrays are known. Although the technology is dependent on which genes or sequences are spotted on the microarrays, the identities of all genes displaying expression differences are known immediately. DNA microarrays are readily applicable to the analysis of gene expression in autopolyploid series such as yeast (Galitski *et al.*, 1999). However, DNA microarrays may not be able to distinguish expression between homologous and/or homoeologous genes in allopolyploids. It requires appropriate experimental design (Fig. 2) and statistical analysis for the detection of transcriptome differences in *Arabidopsis* allotetraploids (Chen *et al.*, 2004) and within and among species (Townsend and Taylor, Chapter 31).

Various microarray platforms and their utilities have been reviewed and discussed (Fodor *et al.*, 1991; Hughes *et al.*, 2001; Johnston, 1998; Lee *et al.*, 2004; Schena *et al.*, 1995, 1998; Singh-Gasson *et al.*, 1999). Here we describe a spotted microarray technology using synthesized long (70-mer) oligos (Lee *et al.*, 2004). We spot about 26,000 oligos designed from all approximately 26,000 annotated genes in the *Arabidopsis* genome available in the National Center of Biological Institutes (NCBI). (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>) and The Institute for Genomic Research (TIGR) (<http://www.tigr.org/tdb/agi/>) (February 2002) (Arabidopsis Genome Initiative 2000). Oligos with a uniform melting temperature ($75 \pm 5^\circ$) are selected within 1000 nucleotides of the 3' end of predicted coding sequences (Lee *et al.*, 2004). Gene names and GenBank accession numbers of the 26,090 70-mer oligos and their corresponding cDNA sequences can be found at <http://www.operon.com/arrays/omad.php>. The slide design, printing, and hybridization of our microarrays are described elsewhere (Lee *et al.*, 2004), except that now each slide is printed with 27,648 features (26,090 oligo-genes plus controls) (Fig. 2E).

Microarray Experimental Design

There are many sources (both biological and technical) of variation that are involved in the microarray experiments; therefore, appropriate experimental design is required for the purpose of accurately estimating biological and technical variation (Churchill, 2002; Kerr and Churchill, 2001; Lee *et al.*, 2004; Quackenbush, 2001). Designing a microarray experiment for the analysis of gene expression is essentially equivalent to planning yield trials in multiple locations and years, as is done by plant breeders. The basic principles are to replicate the experiment so that one can estimate the sources of variation for every (biological or technical) parameter in a statistical model. We employ a dye-swap experimental design for microarray analysis (Fig. 2A–D) (Lee *et al.*, 2004). The advantage of the dye swap is that by definition it includes technical replication, and it allows the estimation of all parameters in a linear model, because every possible combination of factors is observed. In fact when multiple samples are involved in a comparative analysis of gene expression, other methods such as loop and reference designs can be implemented (Churchill, 2002; Kerr and Churchill, 2001). We use a linear model analysis that is equivalent to an analysis of variance (ANOVA) (i.e., the treatments involved in array experiments are typically qualitative). Therefore, each gene can be tested readily for differential expression, and a conclusion drawn that pertains to the statistical significance for that gene. Because so many genes are usually represented on the array, a statistical complication arises due to the very large number of hypothesis tests that are made (commonly referred to as the *multiple comparisons problem* in statistics). To control the type I error across the family of tests made, one can use a variety of multiple comparison corrections (e.g., false-discovery rate [FDR] and family-wise error rate [FWER]) (Hochberg and Benjamini, 1990; Hochberg and Tamhane, 1987).

Using the described dye-swap experimental design, the comparison of gene expression differences between a diploid and isogenic autopolyploid is relatively straightforward (Fig. 2A). However, detecting gene expression differences between an allotetraploid and its two progenitors is complicated. Several comparisons are needed to make the data interpretation meaningful. First, a comparison of gene expression differences between two progenitors determines the levels of divergence between the two parents (Fig. 2B). Second, a comparison between the allotetraploid and midvalue of the two parents (an equal mixture of RNAs from two parents) determines overall changes of gene expression relative to the two parents. However, it does not reveal the parental origin of differentially expressed genes. Silencing or activation of one of the two homoeologous genes may not be detected. Third, to determine the origin of gene expression changes, the allotetraploid is compared with one parent at a time so changes of gene expression in every homoeologous locus may be detected. Because each experiment requires four dye swaps or eight replicates (Chen *et al.*, 2004), a total of 32 slide hybridizations are needed for each experiment involving an allotetraploid and two progenitors (Fig. 2B–D). A microarray image of one slide hybridization between *A. thaliana* and *A. arenosa*, progenitors for *Arabidopsis* allotetraploids, is shown in Fig. 2E. The genes that are highly expressed in *A. thaliana*, *A. arenosa*, or are equally expressed are shown in red, green, and yellow, respectively. In this protocol, we describe microarray manufacturing, probe preparation, slide hybridization and washing, image scanning, and data collection and processing. Statistical models and data analysis tools are described elsewhere (Black and Doerge, 2002; Kerr and Churchill, 2001; Lee *et al.*, 2004).

Oligo Design and Distribution

Each 70-mer oligo is designed from 1000 bp with the 3' end of an annotated gene (Lee *et al.*, 2004). The oligo contains no significant secondary structures, has a melting temperature of $70 \pm 2^\circ$, and is coupled with C₆ amino linker that helps DNA bind to covalent substrate coated on the slides. The database for oligo sequences and corresponding annotated genes are maintained by Qiagen at <http://oligos.qiagen.com/arrays/omad.php>. After synthesis and appropriate

quality control, an equal amount of each oligo (600 pmol) is dispensed into 384-well plates (Genetix, Ltd, Hampshire, UK). Contact Qiagen if the oligos need to be distributed in customized plates. The oligos are lyophilized and stored at -20° .

Array Fabrication

Spotting high-quality slides is the first step toward a successful microarray experiment. We print microarray slides using the OmniGrid Accent microarrayer (GeneMachines, San Carlos, CA) and microspotting Stealth 3 pins (SMP3) or 946 microspotting pins (946MP3) (TeleChem International, Inc., Sunnyvale, CA). The Accent benchtop model is relatively easy to operate and spots 50 slides at a time, while an OmniGrid 300 microarrayer can print 300 slides. We tested various slides coated with different substrates including poly-L-lysine, SuperAmine, aldehyde, phenylene diisocyanate (PDC), and nitrocellulose. The SuperAmine (SMM) (Telechem International, Inc., Sunnyvale, CA) and poly-L-lysine (PLL) (Cel Associates, Inc., Houston, TX) slides are inexpensive and produce the best and uniform spots.

Place the slides carefully onto the microarrayer station. Use an air duster (from any convenience store) to remove debris on the surface of the slides. Adjust the distance between the pins and each slide by calibrating the slide Z-offsets (one motor step is equivalent to $2.5\text{ }\mu\text{m}$) so that each slide is spotted by all 16 pins spontaneously with no missing spots.

Set the speed to “low” with a velocity of 5 cm/s and an acceleration of 30 cm/s^2 for both “dipping” and “spotting.” The pins are blotted (spotting on a clean plastic plate) five times before spotting onto the slides. Between two spotting steps, the pins are cleaned by sonication at 2000 ms followed by three cycles of wash and dry (1000 ms each).

Before printing, the pins should be tested using a blue food dye (McCormick & Company, Inc., Hunt Valley, MD) dissolved in $3\times\text{SSC}$. Each pin should print more than 100 consistent spots. If printed spots are inconsistent, clean the pins using the microspotting pin and printhead cleaning kit (Telechem International). We routinely clean the pins (in addition to a regular clean step during spotting as just mentioned) after each spotting operation or after a long period of storage, to remove surface contaminants such as salts and organic films that may diminish printing performance. Also, clean the pins after about 12 h of continuous spotting (or each of three subsets of spotting; see later discussion).

Fill the ultrasonic bath with 1 liter of 5% microcleaning solution and place pins in a floatable pin rack. Place the rack with pins in the ultrasonic bath and sonicate for 5 min. Rinse the rack and pins exhaustively with tap water. Sonicate for 5 min in ddH_2O , remove pins, and dry. Place clean pins in a printhead and set up for printing.

Centrifuge the plates at 1000 rpm for 2 min to ensure that the oligo pellets fall to the bottom before unsealing the plates.

Make 1 liter of $3\times\text{SSC}$ and attach the bottle to a μFill microplate bulk reagent dispenser (Bio-Tek Instruments, Inc., Winooski, VT). Aliquot $20\text{ }\mu\text{l}$ of $3\times\text{SSC}$ into each well (600 pmol/oligo) of the plates. Seal the plates using a manual sealer (Corning Incorporated, New York, NY), place them on a shaker, and dissolve the oligos overnight. Centrifuge the plates at 1000 rpm before spotting. The final concentration of each oligo is $30\text{ }\mu\text{M}$.

A total of 26,090 oligo samples are distributed in sixty-nine 384-well plates. A control plate contains 16 replicates of 12 negative controls (random 70-mer oligos that do not match any sequences in the databases) and 12 positive controls (70-mer oligos designed from constitutive genes). For spotting, we use 16 pins in 4-by-4 configuration. Each pin prints 576 (or 24 by 24) features per subarray (Fig. 2E). The 16 pins spot three subsets of 16 subarrays or a total of 48

subarrays in each slide, producing a total of 27,648 (16 by 3 by 576) features (Fig. 2E). The 16 pins spot samples from 24 plates (23 sample plates plus the control plate) so that the control plate is spotted in each subset. We set the parameters of low velocity and acceleration for both dipping (picking up the oligos from 384-well plate) and printing (spotting the oligos onto the slide). Each spot is about 100 μm in diameter with 180- by 180- μm spacing between spots. The printing quality is dependent on various mechanical and human factors. We can routinely print high-quality slides (Fig. 2E) using Telechem pins, about 40–65% relative humidity in the printing chamber, stable printing buffer (3 $\times$ SSC), low salt concentration, and PLL or SMM slides. “Double dipping” reduces the number of missing spots during the printing. It takes about 36 h to print 50 slides each with 27,648 features.

To estimate the variation among different slides, we also generate different printing patterns by changing the order of the plates during printing or rotating each plate to 180 degrees so spots in different printing patterns are generated by different pin-sample combinations.

After each spotting operation, we dry the oligo samples in a laminar flow hood, seal the plates using manual dealers (Corning, Inc.), and store at -20° . The oligos can be re-dissolved in 20 μl of ddH₂O overnight before a subsequent printing.

Posttreatment of the Slides

Using a diamond scribe, mark the corners of the printing area that will be covered during hybridization. Arrange 10–15 slides in the same orientation and place them in a slide rack (VWR Scientific Products, West Chester, PA). Hold the slides with the printed side toward open plasticware containing 1 $\times$ SSC heated at 37° . After about 20 s, the surface of the slides becomes foggy. Place the slides on a heating block at 100° and snap-dry the slides for 15 s. Repeat the procedure twice. This procedure allows the DNA spots to “swell” (redistributing the DNA uniformly throughout the spots) but not spread into each other, which results in rounded spot patterns fixed by the heating step.

Place the slides in a Stratalinker UV Crosslinker (Stratagene, La Jolla, CA) and cross-link the DNA to the glass substrate using the setting of 250 mJ.

Prepare the blocking solution: Dissolve 9 g of succinic anhydride in approximately 500 ml of 1-methyl-2-pyrrolidinone (Sigma, St. Louis, MO). Immediately after the last flake of the succinic anhydride is dissolved, add 22.4 ml of 1 M of sodium borate (pH 8.0).

To make 1 M of sodium borate, dissolve 6.2 g of boric acid (Fisher Scientific International, Inc., Fair Lawn, NJ) in a final volume of 100 ml and adjust pH to 8.0 using NaOH.

For blocking the slides, place the slides in a slide rack and quickly plunge the slide rack up and down in the blocking solution for 1 min. Then place the slides in the blocking solution and mix on an orbital shaker at 60 rpm for 15–20 min.

Carefully transfer the slide rack to a slide dish containing ddH₂O and plunge the rack up and down several times to rinse off the blocking solution. Then immerse the slides in ddH₂O for 1 min. (Note: It is critical to remove the residual blocking solution as much as possible to prevent background problems during slide hybridization.)

Transfer the slide rack to plasticware containing 95% ethanol and plunge up and down five to eight times. Dry the slides by centrifugation in a tabletop centrifuge Savant SpeedVac (TeleChem International) at 850 rpm for 5 min. The slides can be used immediately for hybridization or stored in a humidity-controlled Dry Keeper (Sanplatec Corp., Japan) with a relative humidity of 10–20% at RT for up to a year.

Probe Labeling (Direct Dye Incorporation Method)

The procedures for isolating total RNA and purifying mRNA are described in the section “AFLP-cDNA Display,” earlier in this chapter. We use 500 ng of mRNA in each labeling reaction using Cy3- or Cy5-dCTP (Amersham Biosciences). For one dye-swap experiment, we use two sets or four labeling reactions (Fig. 2A). After labeling, the Cy3-dCTP reaction is mixed with the Cy5-dCTP reaction to make one probe. Therefore, two “identical” probes, each containing an equal amount of Cy3- and Cy5-labeled cDNAs, are hybridized with two slides, which constitute one dye-swap experiment.

Primer Annealing

Add 16 μ l of mRNA (~500 ng), 1 μ l of anchored oligo (dT) (2 μ g/ μ l), and 1 μ l of random nonamer (2 μ g/ μ l, Gene Link, Hawthorne, NY) to a 1.5-ml microcentrifuge tube on ice. Mix gently by pipetting. Incubate the reaction at 65° for 5 min. Cool reactions at RT for 10 min to allow the primers and the RNA template to anneal. Centrifuge briefly.

Reverse Transcription

Place reactions on ice and add 6 μ l of reverse-transcriptase buffer (5 $\times$), 3 μ l of DTT (0.1 M), 1 μ l of dNTP (10 mM dATP, dTTP, dGTP, 2.5 mM dCTP), 1 μ l of Cy3- or Cy5-dCTP, and 1 μ l of Superscript II RT (Invitrogen) to a final volume of 30 μ l.

Mix very gently by pipetting and centrifuge briefly. Quickly wrap the tube with aluminum foil and incubate the reactions at 42° for 2 h. From this step on, minimize the direct exposure of the labeling reactions to light.

RNA Degradation

Add 3 μ l of 2.5 M NaOH, mix well by pipetting, and incubate the reactions at 37° for 15 min in the dark (wrap the tube with aluminum foil). Add 15 μ l of 2 M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) to neutralize the solution.

To make 10 ml of 2 M HEPES, dissolve 4.77 g of HEPES (Sigma) in ddH₂O to a final volume of 10 ml, and filter-sterilize the solution using a 0.45- μ m filter (Millipore, Billerica, MA).

Probe Purification

The labeled probes are purified using a PCR purification kit according to the recommendation of the manufacturer (Qiagen). Briefly, add 300 μ l of binding buffer (see earlier discussion) to the reaction and elute twice using 30 μ l of ddH₂O, resulting in 60 μ l. Dry the samples using a Savant SpeedVac (TeleChem International) covered with a black cloth wrap.

Probes can be labeled using two other methods known as (1) amino-allyl dye coupling or postlabeling (Randolph and Waggoner, 1997) (Amersham Biosciences) and (2) 3D labeling with dendritic nucleic acid structure (Nilsen *et al.*, 1997; Stears *et al.*, 2000) (Genisphere, Hatfield, PA). In our experience, the postlabeling method works as well as the direct labeling method. However, for unknown reasons, the dendrimer molecules from the products we tested in 2002 hybridized randomly to some 70-mer oligos (Lee *et al.*, 2004). We have not tested a new product that is claimed to be suitable for labeling probes used in hybridization to spotted oligo microarrays (Genisphere).

RNA Labeling Using CyScribe PostLabeling Kit (RPN5660) (Amersham Biosciences)

The indirect labeling procedure requires two steps. The first step involves incorporation of amino allyl-dUTP (AA-dUTP) during cDNA synthesis. The second step involves conjugation of chemically amino allyl-modified cDNA and cyanine (Cy3 or Cy5) dye. This two-step

incorporation of cyanine dyes presumably improves the labeling efficiency by reducing the bias of direct Cy3- and Cy5-dye incorporation during cDNA synthesis.

We use about 500 ng of mRNA for each indirect labeling reaction according to the recommendations of the manufacturer (Amersham Biosciences). Synthesize the first-strand cDNA using AA-dUTP and Superscript II RT (Invitrogen). The Superscript II RT works better than the Cyscript RT provided in the kit. Incorporate the Cy3 or Cy5 dye to amino allyl-modified cDNA through the reaction of the amino allyl group with the monofunctional *N*-hydroxysuccinimide (NHS)-ester Cy dye (Amersham Biosciences) according to the manufacturer's recommendations. The labeled cDNA is then purified using a PCR purification kit (Qiagen) as described earlier.

Hybridization

Resuspend the dried probe in 56 μ l of ddH₂O and denature at 95° for 3 min. Add 12 μ l of 20× SSC, 2 μ l of 10% SDS, and 8 μ l of 10% BSA to the probe, and mix well in a final volume of about 80 μ l.

Place the posttreated microarray slide on a heating plate (at 40–50°), add the probe as one drop of liquid to an area that is 5 mm from the marked edge of the printed area, and slowly lay down a 22-by-26 LifterSlip (Erie Scientific Company, Portsmouth, NH) over the hybridization solution so that the probe is evenly distributed without bubbles.

Alternatively, lay down the LifterSlip and add the 2/3 of labeled probe very slowly from one side of the coverslip. The solution flows to completely cover the hybridization area so no bubbles are formed. Add the rest of the probe from the opposite side of coverslip. (Do not let the pipette tip move the LifterSlip.)

Place the slide in a sealed hybridization cassette (TeleChem International) and add 20 μ l of 3× SSC to both ends of the slide in the cassette. Place the hybridization cassette in a hybridization oven at 65° for 12–16 h. We normally hybridize four or eight slides each time. The hybridization and washing should be performed in dark conditions whenever possible.

Remove each slide from the hybridization cassette without disturbing the coverslip and place the slides in a slide rack. Place the slide rack in a washing station (HTW) (Telechem International) containing low-stringency washing buffer (2× SSC, 0.2% SDS). The coverslip gradually slips off the slide into washing buffer in less than 1 min. Gently remove the coverslip and continue washing slides for 4 min at RT.

Wash the slides in 0.2× SSC buffers for 2 min at RT and repeat once. Wash the slides at high-stringency in another washing station containing 0.05× SSC for 4 min at RT. Remove the slide rack and plunge the slides up and down several times in the washing station containing fresh 0.05× SSC buffer. Dry the slides immediately by centrifugation in an Allegra X-12 tabletop centrifuge (Beckman Coulter, Inc., Fullerton, CA) at 850 rpm for 10 min. The slides then are assembled in a slide rack and quickly placed in a dark container.

Microarray Scanning (Parameters) and Data Analysis

Take one slide at a time, carefully place it in the scanning chamber of a GenePix 4000B scanner (Axon Instruments, Inc., Union City, CA), and gently close the lid. The images can be scanned using “preview scan” function to estimate the relative intensities of the two-color images. The red and green signals need to be balanced during array scanning by adjusting the photomultiplier tube (PMT, maximum 1000) gain and 100% laser power for each wavelength. We normally set the power to 100% and the PMT gain to 900–990 for the 635-nm (red) laser and 800–900 for the 532-nm (green) laser. Each microarray spot is about 100 μ m in diameter.

When the slide is scanned using the setting of 10- μ m pixel size, each spot contains about 78 pixels, which is sufficient for an accurate intensity measurement using GenePix Pro 5.0 software. A pixel with an intensity of 65,535 or more is saturated and seen as a “white spot” on the slide. The minimum intensity that can be detected is 500. While scanning, we adjust the PMT gain in each channel using a “count ratio” until the ratio is about 1.0. After “preview scan,” we select the area and scan it again using the “data scan” function.

After the scanning is completed, we save the image as a multi-image TIFF file that can be exported as a JPEG file. The PMT values and probe/hybridization information for each file are recorded according to the requirement for maintaining minimum information about microarray experiments (MIAME) (Brazma *et al.*, 2001). For the quantification of each hybridized feature, the “deconvolution blocks” are generated by the GeneMachines microarrayer using “generate grid” (384 by 72), which is exported to the GenePix Pro 5.0 as a GAL file. In the GenePix Pro 5.0, we use “loading array list” to overlay the “grid” to the scanned images. We then adjust the grid to fit the features automatically, followed by manual corrections so that the quantification area is localized in the center of every feature. The intensities of each feature at 532 and 635 nm are generated in a GPR file for further analysis. GenePix Pro 5.0 has a few analysis functions to display scatter plots and intensity ratios.

Data Normalization and Analysis

The GPR files generated from each experiment are imported from GenePix Pro 5.0 and analyzed using both Matlab (MathWorks, Inc., Natick, MA) (Ihaka and Gentleman 1996) and custom statistical programs written with the statistical package known as R (<http://www.r-project.org>). Data normalization can be used to correct some technical variations involved in microarray experiments, which includes unequal amounts of RNA samples, different spotting pins, variable surface on the slides, different labeling and detection efficiencies between the cyanine dyes used, and systematic biases in the measured expression levels (Quackenbush, 2002). Normalization of microarray data is equivalent to the use of a reference or control to estimate the expression levels of the genes of interest relative to the control. Typically, data normalization is not necessary for the ANOVA model (Lee *et al.*, 2004). The data may be collected directly from scanned images and transformed using $\log_2(\text{ratio})$ total intensities. However, in many microarray experiments, local variation is commonly observed because of factors such as unequal hybridization, biased dye incorporation, and block (or pin)-to-block variation. A robust locally-weighted linear regression (lowess) (Cleveland, 1979) is recommended for two-color microarray experiments. This normalization process allows the correction for the bias caused by nonlinear rates of dye incorporation and other systematic technical factors such as pins. Lowess normalization can be applied either globally or locally. Local normalization has the advantage of correcting systematic spatial variation that is often observed in large array experiments. In our experiments, we use the local (pin-based) lowess normalization and analyze data using the ANOVA model (Chen *et al.*, 2004). The multiple comparison tests using a per-gene variance and a common gene variance are employed to select differentially expressed genes. Holm’s FWER and/or FDR (Benjamini and Hochberg, 1995; Hochberg and Tamhane, 1987) are applied to control multiple comparison testing errors at a significant level $\alpha = 0.05$.

Commercial software packages such as GeneSpring (Silicon Genetics, Redwood City, CA) and Spotfire (Spotfire, U.S., Somerville, MA) are available for the data normalization, quality control, and data analysis. A caveat for using commercial software is that underlying statistical models (how to calculate variance, multiple comparisons corrections, etc.) are not clear. Biologists should consult with a statistician/bioinformatician to be sure that their experimental design is compatible with the underlying models in these computer programs.

The differentially expressed genes detected (Fig. 2) are confirmed and analyzed using other methods, such as quantitative RT-PCR, RT-PCR and CAPS, and SSCP (Adams *et al.*, 2003; Lee and Chen, 2001; Lee *et al.*, 2004).

Notes

1. Current printing patterns do not contain replications of features within a slide. If the statistical model requires replications, every feature can be duplicated in each slide. However, spotting about 60,000 features in one slide may decrease spacing between spots and resolution in subsequent hybridization and detection. Alternatively, a set of 26,000 oligos can be printed on two slides, although this increases variation between slides.
2. Slide hybridization using a LifterSlip cover is one of the best choices for us, because the printed area occupies the maximum amount of space on the slide and leaves little room for sealers that are used in many automated systems. It is recommended that before using probe-containing hybridization solutions, one should practice several times using the LifterSlip to cover hybridization solutions without probes on a regular slide until no bubbles are formed.
3. We recommend using direct or indirect labeling methods for probe preparation. The direct incorporation method is routinely used because the Superscript II RT (Invitrogen) has a similar efficiency for Cy3- and Cy5-dCTP incorporation. However, when mRNA is limiting (e.g., from cotton fiber tissues in early stages), the Amino Allyl MessageAmp aRNA kit (Ambion, Austin, TX) can be used to label probes for microarray hybridization. Only 5 μ g of total RNA is required for each labeling reaction (Lee and Chen, unpublished data).
4. For genomic DNA analysis, we label genomic DNA using a modified protocol (Pollack *et al.*, 1999). For each labeling, 1 μ g of genomic DNA is digested by *Mse* I and purified using a QIAquick PCR kit, which is then random-primer labeled using a Bioprime labeling kit (Gibco BRL). The labeling reaction includes 1 μ g of *Mse* I-digested genomic DNA, dATP, dGTP, and dTTP (120 μ M each), dCTP (90 μ M), and Cy3-dCTP or Cy5-dCTP (30 μ M) in 50 μ l of reaction. The labeled probe is purified as described earlier. The same hybridization and washing procedures are used except that 20 μ g of human cot-1 DNA and 100 μ g of tRNA (Invitrogen) are included in about 80 μ l of hybridization solution.
5. We have shown that spotted 70-mer oligo microarrays can be used to hybridize homoeologous and orthologous genes in *Arabidopsis* and related species including *Brassica* (Lee *et al.*, 2004).
6. Although the spotted oligo microarray has many advantages (Lee *et al.*, 2004), the oligo design is dependent on accuracy of gene annotation. Therefore, gene indices should be updated periodically according to the new release of annotated databases from research groups such as TAIR (<http://arabidopsis.org/>) and TIGR (<http://www.tigr.org/>).
7. PLL slides need to be freshly made. High levels of hybridization background are often associated with old PLL slides. We prefer SMM to PLL slides.

Acknowledgements

We thank James A. Birchler, Robert Martienssen, Thomas C. Osborn, and J. Chris Pires for comments on the manuscript and insightful discussions leading to the development of spotted oligo gene microarrays in *Arabidopsis*. The research is supported by a grant from the National Science Foundation (DBI0077774). Work in the Chen laboratory is supported in part by the National Institutes of Health (GM067015). The opinions expressed are those of the authors

and do not reflect the official policy of the National Institutes of Health, the National Science Foundation, or the U.S. government.

References

- Adams KL, Cronn R, Percifield R, Wendel JF. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci USA* 2003;100:4649–4654. [PubMed: 12665616]
- Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000;408:796–815. [PubMed: 11130711]
- Bachem CW, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeau M, Visser RG. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. *Plant J* 1996;9:745–753. [PubMed: 8653120]
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B* 1995;57:289–300.
- Black MA, Doerge RW. Calculation of the minimum number of replicate spots required for detection of significant gene expression fold change in microarray experiment. *Bioinformatics* 2002;18:1609–1616. [PubMed: 12490445]
- Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–371. [PubMed: 11726920]
- Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 2000;18:630–634. [PubMed: 10835600]
- Chen ZJ, Wang J, Tian L, Lee HS, Wang JJ, Chen M, Lee JJ, Josefsson C, Madlung A, Watson B, Lippman Z, Vaughn M, Pires JC, Colot V, Doerge RW, Martienssen RA, Comai L, Osborn TC. The development of an *Arabidopsis* model system for genome-wide analysis of polyploidy effects. *Biol J Linn Soc* 2004;82:689–700.
- Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 2002;32:490–495. [PubMed: 12454643]
- Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979;74:829–836.
- Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* 2000;12:1551–1568. [PubMed: 11006331]
- Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991;251:767–773. [PubMed: 1990438]
- Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR. Ploidy regulation of gene expression. *Science* 1999;285:251–254. [PubMed: 10398601]
- Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med* 1990;9:811–818. [PubMed: 2218183]
- Hochberg, Y.; Tamhane, AC. “Multiple Comparison Procedures”. Wiley; New York: 1987.
- Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephanian SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 2001;19:342–347. [PubMed: 11283592]
- Ihaka R, Gentleman R. A language for data analysis and graphics. *J Comput Graphical Statist* 1996;5:299–314.

- Johnston M. Gene chips: Array of hope for understanding gene regulation. *Curr Biol* 1998;8:R171–R174. [PubMed: 9501061]
- Kerr MK, Churchill GA. Experimental design for gene expression microarrays. *Biostatistics* 2001;2:183–201. [PubMed: 12933549]
- Klein PE, Klein RR, Cartinhour SW, Ulanich PE, Dong J, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE. A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. *Genome Res* 2000;10:789–807. [PubMed: 10854411]
- Lee HS, Chen ZJ. Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proc Natl Acad Sci USA* 2001;98:6753–6758. [PubMed: 11371624]
- Lee HS, Wang J, Tian L, Jiang H, Black MA, Madlung A, Watson B, Lukens L, Pires JC, Wang JJ, Comai L, Osborn TC, Doerge RW, Chen ZJ. Sensitivity of 70-mer oligonucleotides and cDNAs for microarray analysis of gene expression in *Arabidopsis* and *Brassica*. *Plant Biotechnol J* 2004;2:45–57. [PubMed: 17166142]
- Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 1992;257:967–971. [PubMed: 1354393]
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999;21:20–24. [PubMed: 9915496]
- Liu B, Wendel J. Non-Mendelian phenomenon in allopolyploid genome evolution. *Curr Genomics* 2002;3:489–505.
- Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiol* 2002;129:733–746. [PubMed: 12068115]
- Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ. Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). *Theor Applied Genet* 2004;108:280–291. [PubMed: 14513220]
- Nilsen TW, Grayzel J, Prenskey W. Dendritic nucleic acid structures. *J Theor Biol* 1997;187:273–284. [PubMed: 9237897]
- Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V, Martienssen RA. Understanding mechanisms of novel gene expression in polyploids. *Trends Genet* 2003;19:141–147. [PubMed: 12615008]
- Pikaard CS. Genomic change and gene silencing in polyploids. *Trends Genet* 2001;17:675–677. [PubMed: 11718903]
- Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999;23:41–46. [PubMed: 10471496]
- Quackenbush J. Computational analysis of microarray data. *Nat Rev Genet* 2001;2:418–427. [PubMed: 11389458]
- Quackenbush J. Microarray data normalization and transformation. *Nat Genet* 2002;32:496–501. [PubMed: 12454644]
- Randolph JB, Waggoner AS. Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes. *Nucleic Acids Res* 1997;25:2923–2929. [PubMed: 9207044]
- Sambrook J, Fritsch EF, Maniatis T. “Molecular Cloning: A Laboratory Manual”. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 1989.
- Schena M, Heller RA, Thieriault TP, Konrad K, Lachenmeier E, David RW. Microarrays: Biotechnology’s discovery platform for functional genomics. *Trends Biotechnol* 1998;16:301–306. [PubMed: 9675914]
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–470. [PubMed: 7569999]
- Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR, Cerrina F. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol* 1999;17:974–978. [PubMed: 10504697]
- Stears RL, Getts RC, Gullans SR. A novel, sensitive detection system for high-density microarrays using dendrimer technology. *Physiol Genomics* 2000;3:93–99. [PubMed: 11015604]

- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 1995;270:484–487. [PubMed: 7570003]
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res* 1995;23:4407–4414. [PubMed: 7501463]
- Wang J, Tian L, Madlung A, Lee HS, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* 2004;167:1961–1973. [PubMed: 15342533]
- Welsh J, Chada K, Dalal SS, Cheng R, Ralph D, McClelland M. Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res* 1992;20:4965–4970. [PubMed: 1383934]

Fig 1.

Amplified fragment length polymorphic complementary DNA (AFLP-cDNA) display analysis of gene expression. (A) A simple diagram of the AFLP-cDNA display procedures including messenger RNA (mRNA) isolation, cDNA synthesis, restriction digestion, ligation, preselective and selective amplification, and gel electrophoresis. The detailed protocols are described in the text. Only a few hypothetical fragments are shown to illustrate how *EcoR* I (GAATTC)- and *Mse* I (TTAA)-digested fragments are ligated and amplified during the analysis. (B) AFLP-cDNA display analysis using leaves of *Arabidopsis thaliana* diploid (At2), autotetraploid (At4), *Arabidopsis arenosa* (Aa), and generations 2–5 (S2–S5) of new allotetraploids, and a natural *Arabidopsis suecica* (As) line (Wang *et al.*, 2004). The primers used in selective amplification are E-TA and M-CTG. Open and filled squares indicate that the genes are up-regulated and down-regulated, respectively, in the isogenic autotetraploid (At4). Arrows indicate gene silencing or activation (filled) and random changes in gene expression (open) in different generations of the allotetraploids, respectively. Negative controls using mRNA samples without reverse transcriptase in the reaction (Lee and Chen, 2001; Wang *et al.*, 2004) are not shown because of space limitations.

Fig 2.

Microarray experimental design and analysis for detecting gene expression changes in autotetraploids and allotetraploids. (A) A dye-swap experiment using messenger RNA (mRNA) from a diploid and an isogenic autotetraploid to study changes of gene expression in response to ploidy increase. Four labeling reactions using Cy3- and Cy5-dCTP are prepared and mixed reciprocally (e.g., 1 + 3 and 2 + 4) to hybridize two slides in one dye-swap experiment. Four dye-swap experiments (or eight slides) are used in each comparison. (B–D) To detect genes that are differentially expressed in an allotetraploid, four sets of dye-swap experiments are performed for each comparison between the two parents (B), between the allotetraploid and two progenitors (C), or between the allotetraploid and each progenitor (D). Results from the experiments in (B) and (C) determine the overall changes of gene expression in the allotetraploid relative to the progenitors but may underestimate the number of homoeologous genes that are silenced (see text). (E) A hybridization image shows 27,648 features in the *Arabidopsis* spotted 70-mer oligogene microarray, as previously described (Chen *et al.*, 2004). The slide is hybridized to complementary DNA (cDNA) prepared from *Arabidopsis thaliana* and *Arabidopsis arenosa*. The red, green, and yellow-white features in the subarray indicate low, high, and equal levels of gene expression, respectively, in *A. thaliana*. The data obtained from four sets of dye-swap experiments in each comparison are analyzed using a linear model (Lee *et al.*, 2004) to detect differentially expressed genes at a significance level ($\alpha = 0.05$, see text).