The development of an *Arabidopsis* model system for genome-wide analysis of polyploidy effects

Z. JEFFREY CHEN¹,*, JIANLIN WANG¹, LU TIAN¹, HYEON-SE LEE¹, JIYUAN J. WANG¹, MENG CHEN¹, JINSUK J. LEE¹, CAROLINE JOSEFSSON², ANDREAS MADLUNG²,†, BRIAN WATSON², ZACH LIPPMAN³, MATT VAUGHN³, J. CHRIS PIRES⁴, VINCENT COLOT³,‡, R. W. DOERGE⁵, ROBERT A. MARTIENSSEN³, LUCA COMAI², and THOMAS C. OSBORN⁴

¹Intercollegiate Program in Genetics and Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843-2474, USA

²Department of Botany, University of Washington, Seattle, WA 98195, USA

³Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

⁴Department of Agronomy, University of Wisconsin, Madison, WI 53706, USA

⁵Departments of Statistics, Agronomy, and Computational Genomics, Purdue University, West Lafayette, IN 47907, USA

Abstract

*Arabidopsis* is a model system not only for studying numerous aspects of plant biology, but also for understanding mechanisms of the rapid evolutionary process associated with genome duplication and polyploidization. Although in animals interspecific hybrids are often sterile and aneuploids are related to disease syndromes, both *Arabidopsis* autopolyploids and allopolyploids occur in nature and can be readily formed in the laboratory, providing an attractive system for comparing changes in gene expression and genome structure among relatively ‘young’ and ‘established’ or ‘ancient’ polyploids. Powerful reverse and forward genetics in *Arabidopsis* offer an exceptional means by which regulatory mechanisms of gene and genome duplication may be revealed. Moreover, the *Arabidopsis* genome is completely sequenced; both coding and non-coding sequences are available.

We have developed spotted oligo-gene and chromosome microarrays using the complete *Arabidopsis* genome sequence. The oligo-gene microarray consists of ~26 000 70-mer oligonucleotides that are designed from all annotated genes in *Arabidopsis*, and the chromosome microarray contains 1 kb genomic tiling fragments amplified from a chromosomal region or the complete sequence of chromosome 4. We have demonstrated the utility of microarrays for genome-wide analysis of changes in gene expression, genome organization and chromatin structure in *Arabidopsis* polyploids and related species.

ADDITIONAL KEYWORDS

Brassicaceae; chromatin; epigenetics; gene silencing; genome dosage; genome duplication; microarray; RNAi; transcriptome; transposons

*Corresponding author. E-mail: zjchen@tamu.edu.
†Current address: Department of Biology, University of Puget Sound, Tacoma, WA 98416, USA
‡Current address: Unité de Recherche en Génomique Végétale, Evry, Paris, France.
INTRODUCTION

Since the first polyploidy conference in 1979 (Lewis, 1980), polyploidy research has undergone a renaissance. Gene activation and silencing and genomic changes have been detected in both natural and newly formed polyploids in a variety of plant species (Song et al., 1995; Feldman et al., 1997; Comai et al., 2000; Lee & Chen, 2001; Ozkan, Levy & Feldman, 2001; Kashkush, Feldman & Levy, 2002; Madlung et al., 2002; Adams et al., 2003; Kashkush, Feldman & Levy, 2003 and Doyle et al., 2004; Kovarik et al., 2004; Lukens et al., 2004; Pires et al., 2004 – all this issue). Results from many of these studies have been documented and perspectives reviewed (Soltis & Soltis, 1995; Matzke, Scheid & Matzke, 1999; Pikaard, 1999; Comai, 2000; Pikaard, 2000a; Wendel, 2000; Liu & Wendel, 2002; Osborn et al., 2003 and Adams & Wendel, 2004; Levy & Feldman, 2004 – both this issue). However, the approach used in many of these studies involved only subsets of genes and/or DNA markers; little is known about polyploidy effects on the entire sets of duplicate or homoeologous (partially homologous) genomes. Here we describe DNA microarrays and several other genome-wide approaches being developed in the model plant Arabidopsis for analysing changes in gene expression, genome organization and chromatin structure during polyploid formation and evolution.

ARABIDOPSIS AS A MODEL SYSTEM FOR STUDYING MOLECULAR MECHANISMS OF GENOME DUPLICATION

Arabidopsis has been recognized as a model system because of its small genome, rapid life cycle, high transformation efficiency, completely sequenced genome, and powerful reverse and forward genetics (Meyerowitz & Pruitt, 1985; Meyerowitz, 1989; Arabidopsis Genome Initiative, 2000; Somerville & Koornneef, 2002). Arabidopsis belongs to the Brassicaceae or Crucifer family, which includes the related genera Arabis, Brassica and Cardamine (Price, Palmer & Al-Shehbaz, 1994; Yang et al., 1999; Koch, Haubold & Mitchell-Olds, 2000; Koch, Haubold & Mitchell-Olds, 2001; Hall, Fiebig & Preuss, 2002). Like many flowering plants, genome duplication and polyploidization are common in the genus Arabidopsis and its related species. Estimates indicate that the ‘diploid’ A. thaliana underwent at least one event of whole genome duplication ~24–40 Mya (Vision, Brown & Tanksley, 2000; Blanc, Hokamp & Wolfe, 2003) before the divergence of Arabidopsis–Brassica (~12–20 Mya) and A. thaliana–A. arenosa (~5.8 Mya) (Koch et al., 2001). Furthermore, hybridization between related species occurs in Arabidopsis, giving rise to allopolyploids. Indeed, some Arabidopsis species are either autopolyploid or allopolyploid.

Arabidopsis autopolyploids occur naturally, presumably through fertilization of unreduced male and female gametes. Naturally occurring A. thaliana (ecotype Columbia) autotetraploids have been described (Weiss & Maluszynska, 2000) (G. Redei, CS3151 & 3245, Arabidopsis Biological Resource Center) and A. arenosa is a natural and self-incompatible autotetraploid (Malmgren, 1982; Lind-Hallden, Hallden & Sall, 2002). In the laboratory, autotetraploid A. thaliana can be created by germinating inbred diploid seeds in the presence of 0.5–1% colchicine (Fig. 1). The ploidy levels of individual plants are verified by flow cytometry and chromosome counting in cells from flower buds or root tips (Comai et al., 2000). Triploids are generated by crossing diploids with tetraploids. As such, a ploidy series can be made from a single inbred line in the same genetic background or ecotype. These lines are excellent materials for polyploidy research because one may compare changes of gene expression and genome organization in response to alteration of genome dosage (e.g. diploid, triploid and tetraploid) in exactly the same genetic background; we refer to these as ‘matched’ autopolyploids. Moreover, the effects of various ecotypes (e.g. Columbia, Landsberg erecta or Zurich) on ploidy-induced gene regulation can also be investigated.
Naturally occurring *Arabidopsis* allotetraploids are likely to be as common as natural autopolyploids. *A. suecica* is a natural and selfing allotetraploid native to northern Europe (Hylander, 1957; Love, 1961; O’Kane, Schaal & Al-Shehbaz, 1995; Lind-Hallden et al., 2002; Säll et al., 2003). According to data obtained from chloroplast DNA and Rubisco protein analyses, *A. thaliana* is the maternal parent of the allotetraploid *A. suecica* (Mummenhoff & Hurka, 1994; Price et al., 1994). Thus, it is reasonable to assume that the natural *A. suecica* is produced by pollinating *A. thaliana* (e.g. unreduced 2n female gametes) with 2x pollen from *A. arenosa*. The 26 chromosomes of *A. suecica* (2n = 4x = 26) comprise two sets: ten from the diploid selfing species *A. thaliana* (2n = 2x = 10) and 16 from the autotetraploid outcrossing species *A. arenosa* (2n = 4x = 32) (Fig. 2A). The parental origin of the *A. suecica* chromosomes has been determined by DNA sequence analyses and by in situ hybridization using repeated DNA sequences of either *A. thaliana* or *A. arenosa*, which paint ten and 16 centromeres, respectively (Price et al., 1994; Kamm et al., 1995; O’Kane et al., 1995) (Fig. 2C). Although they are very closely related species, these two taxa exhibit 5–8% nucleotide divergence in protein-coding sequences (Hanfstingl et al., 1994; Henikoff & Comai, 1998; Lee & Chen, 2001) and 30–40% divergence in the 180-bp centromeric repeats (Kamm et al., 1995). The species most closely related to *A. thaliana* are *A. arenosa*, *A. neglecta* and *A. lyrata* (O’Kane et al., 1995). Attempts have been made to synthesize new hybrids between *A. thaliana* and *A. lyrata*, but many hybrid lines are sterile and can only be maintained by backcrossing to one of the parent lines (Nasrallah et al., 2000). New allotetraploids between *A. thaliana* and *A. arenosa*, by contrast, can be readily made in the laboratory (Comai et al., 2000; Bushell, Spielman & Scott, 2003).

As part of the consortium project on the ‘functional genomics of plant polyploids’ (http://polyploid.agronomy.wisc.edu), we have produced a series of *Arabidopsis* autopolyploid lines using the method described above, as well as multiple independent lineages of *A. suecica*-like allotetraploids (Fig. 2B). Synthetic *A. suecica*-like lines are made by crossing an autotetraploid *A. thaliana* to the autotetraploid *A. arenosa* (Chen, Comai & Pikaard, 1998; Comai et al., 2000). The resulting F1S have two sets of each parental genome as determined by in situ hybridization (Fig. 2C). F1S and following generations are self-compatible (although the parent *A. arenosa* is self-incompatible), but vary in fertility and viability. The resynthesized lines have been inbred for four additional generations (S2 to S5) and are presumably homozygous for the majority of loci (see diagram in Fig. 2A), although they show some degree of phenotypic variation (Fig. 2B), which may be due, at least in part, to heterozygosity of *A. arenosa*, an outcrossing autotetraploid. Furthermore, gene activation, silencing or rapid genomic changes may account for this phenotypic variation (Osborn et al., 2003). The new and natural strains of *A. suecica* behave as a disomic polyploid or functional diploid (pairing occurs only between homologous chromosomes) during meiosis (Fig. 2C). To study natural variation and population changes, we have collected a variety of polyploid ecotypes. For example, 15 *A. suecica* ecotypes have been collected from various locations in Europe. Five *A. arenosa* ecotypes are available; four have white flowers and one has pink flowers. These lines are extremely valuable for comparative studies on gene expression changes between newly formed *A. suecica*-like lines and old ‘established’ ecotypes, among different generations and among different lineages. Furthermore, dominant negative mutants of chromatin modifying genes are available in *Arabidopsis* diploids (http://www.chromdb.org/) using RNA interference (Chuang & Meyerowitz, 2000; Waterhouse, Wang & Lough, 2001) and can be generated in *Arabidopsis* polyploids (Lawrence & Pikaard, 2003; Wang et al., 2004). T-DNA insertional knockouts for almost every *Arabidopsis* gene have been generated (http://signal.salk.edu/tabout.html). These lines will be useful for studying the role of chromatin in controlling the function of duplicate genes in polyploids as previously reviewed (Osborn et al., 2003).

*Arabidopsis suecica* is an attractive model for studying mechanisms of genome change in allotetraploids because the genome of *A. thaliana* ecotype Columbia has been sequenced. A
recent milestone in this research is the finding of gene activation and silencing in new and natural allotetraploid *A. suecica* (Comai *et al.*, 2000; Lee & Chen, 2001; Madlung *et al.*, 2002). The results indicate that many genes encoding various proteins, mobile elements and non-coding RNAs are silenced. To our best estimate, about ~1.5–2.5% of genes are silenced epigenetically in natural and synthetic *A. suecica* lines. Apart from the rDNA loci in nucleolar organizer regions (NORs), the silenced genes do not appear to be organized in chromatin domains because the genes adjacent to them remain co-expressed (Lee & Chen, 2001; Lewis & Pikaard, 2001). Furthermore, silencing acts on genes of either parental origin, and so does not appear to be selective. Silencing of these genes is maintained by DNA methylation and their expression is reactivated by 5-aza-2′-deoxycytidine, a DNA methyltransferase inhibitor. In nucleolar dominance, silencing of the rRNA genes is maintained by DNA methylation and histone modification (Pikaard, 2000b). Blocking DNA methylation or histone deacetylation or both was shown to reactivate the silenced genes (Chen & Pikaard, 1997), which provides direct evidence that both DNA and histone modifications are responsible for epigenetic regulation of duplicate genes in polyploids. Like the DNA code, combinatorial histone modifications can be considered to be a code and transmitted through meiosis (Jenuwein & Allis, 2001).

**DEVELOPMENT OF *ARABIDOPSIS* Oligo-Gene Microarray for Genome-Wide Analysis in Polyploids**

Previous studies on genome-wide expression changes in polyploids employed a PCR-based technique, AFLP-cDNA display (Comai *et al.*, 2000; Lee & Chen, 2001; Madlung *et al.*, 2002). The advantage of using AFLP-cDNA display is that the origin of the expressed homoeologous transcripts may be determined by comparing the polymorphic cDNA fragments between the allotetraploids and their progenitors. However, the efficiency of AFLP-cDNA display is limited by post hoc cloning and sequencing of individual cDNA fragments. Other assays similar to AFLP-cDNA display such as serial analysis of gene expression (SAGE) (Velculescu *et al.*, 1995; Velculescu *et al.*, 1997) 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 biology procedures involving adaptor ligation, PCR amplification, and sequencing or cloning various signatures.

DNA microarrays are a recently developed technology based on hybridization of labelled DNA, cDNA or cRNA to specific DNA sequences (features) that are fixed on to a substrate. They offer some advantages over AFLP-cDNA display and other methods for transcription analysis because they can be designed to interrogate all gene sequences simultaneously. Because the sequences of the features on the microarrays are known, the identities of all genes displaying expression differences are known immediately. One disadvantage is that they may not be able to distinguish between expression of homologous and/or homoeologous genes. DNA microarrays have been applied to numerous transcriptome studies, including a study on the effects of autopolyploidy on gene expression in yeast (Galitski *et al.*, 1999).

DNA microarrays can be classified into predesigned chips and spotted microarrays based on how the DNA substrates are applied to the slide (Schena *et al.*, 1998). Predesigned chips (e.g. Affymetrix chips) employ direct synthesis of short oligonucleotides (15–25-mers) on a glass substrate using photolithography (Fodor *et al.*, 1991; Lipshutz *et al.*, 1999) or a micromirror device (Singh-Gasson *et al.*, 1999). In spotted microarrays, a spotting robot (Schena *et al.*, 1995) or inkjet device (Hughes *et al.*, 2001) is used to deposit genomic DNA, cDNA amplicons or synthesized long oligonucleotides (50–70-mers) on to a glass slide. Compared with the predesigned chips, the spotted microarrays are relatively inexpensive and offer more flexibility in experimental design and data analysis. Compared with cDNA amplicons, the spotted oligo-microarrays have uniform length, quantity and quality, and low error rates, because quality controls for synthesizing and tracking oligos are usually implemented by automation. In cDNA...
microarrays a high error rate is often associated with clone contamination or amplicon tracking (Knight, 2001). Moreover, spotted oligo-microarrays can include every predicted gene from a fully sequenced genome whereas cDNA microarrays represent only a portion of the genes that have been isolated as expressed sequence tags (ESTs). For organisms whose genome sequences are incomplete, the spotted oligo-microarrays are flexible so that the microarrays can be easily expanded to include new sets of oligo-genes (based on new genomic and EST sequences) using the same array format. The sensitivity of these microarrays in detecting gene expression decreases from cDNA, to long oligos, to short oligos but the specificity of detection efficiency increases from cDNA amplicons to short oligos (Dudley et al., 2002; Relogio et al., 2002). Thus, both spotted oligos and predesigned chips have the potential to detect splicing variants and members of multigene families. However, the specificity of short oligos (15–25-mers) is probably too high, preventing their use in expression studies of related species.

We are utilizing spotted oligo-gene microarrays containing all ~26 000 annotated genes in the Arabidopsis genome for gene expression analysis of Arabidopsis and Brassica polyploids. The oligos (70-mers) were designed for 26 090 Arabidopsis genes using multiple BLAST searches against annotation databases 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 ± 5 °C) were selected within 1000 nucleotides of the 3’ end of predicted coding sequences. Each oligo was coupled with amino-linkers at the 5’-C6 (Operon/Qiagen, California). Gene names and GenBank accession numbers of the 26 090 70-mer oligos and their corresponding cDNA sequences can be found in 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. 3). Sets of 70-mer oligos have been developed using information from all annotated genes of complete genome sequences for several model organisms, including yeast, Drosophila, human, mouse and Arabidopsis.

**ARABIDOPSIS CHROMOSOME 4 MICROARRAY**

Most microarray designs are based on annotated genes, and most genes are annotated based on their expression. However, silent genes and transposable elements are not, by definition, represented in cDNA libraries but some of these genes have been shown to be activated in polyploids (Comai et al., 2000; Madlung et al., 2002; Kashkush et al., 2003). Furthermore, non-coding sequences, such as small nuclear RNA and defective transposons, are not represented on these arrays. Therefore, to complement the oligo-gene microarrays described above, we have designed a chromosomal microarray in which successive 1 kb fragments of chromosome 4 are spotted on to glass slides. Fragments are amplified using custom primer sets, which are designed taking the repetitive content of the fragment into account. In a pilot project, the chromosome microarray containing the knob region (~2 Mb) of chromosome 4 (Fransz et al., 2000) has been developed. Details of the primer design and analysis will be published elsewhere (Lipmann et al., 2004).

**MICROARRAY DATA ANALYSIS USING LINEAR MODELS**

Various factors affect the detection efficiency and sensitivity of microarray experiments. Mechanical and technical limitations, such as defective pins, irregular printing patterns and variable hybridization intensities, may cause systematic errors of microarray data. Many of these limitations have been extensively discussed (Hegde et al., 2000; Finkelstein et al., 2002) and procedures for maintaining minimum information about microarray experiments (MIAME) (Brazma et al., 2001) have been widely adopted by microarray research groups.
The detection sensitivity for significant fold changes of gene expression must be determined within the context of the design used for an experiment. We employ a dye-swap design (Black, 2002) to compare the sources of variation, both technical and biological, to be accounted for via a linear model (Kerr & Churchill, 2001; Black & Doerge, 2002). When multiple treatments are involved in each experiment, loop and reference designs (Quackenbush, 2001; Yang & Speed, 2002) are applicable. The dye-swap is a Latin Square statistical design and therefore the experiment is balanced (i.e. every combination of factors is observed) (Table 1). Using a linear model, we can estimate the sources of variation and test the significance of variation due to the treatment. Moreover, inherent to the balanced design is the ability to estimate all factors using analysis of variance (ANOVA). Typically, when the assumptions for the equal variance are violated, it is possible to accommodate this violation with per-gene variance calculations and multiple comparison adjustments (Hochberg & Benjamini, 1990; Benjamini & Hochberg, 1995). The power of the dye-swap experimental design and linear model analysis has been demonstrated in recent studies in which the detection sensitivities for significant genes are evaluated on a per-gene basis with statistical significance and confidence (Black & Doerge, 2002; Lee et al., 2004). Appropriate experimental designs (described below) and statistical analyses are exceedingly important in microarray experiments because many genes may show small fold-differences in response to polyploidy changes or other biological treatments.

APPLICATION OF MICROARRAYS TO GENOME-WIDE ANALYSIS IN ARABIDOPSIS POLYPLOIDS AND RELATED SPECIES

The two types of microarrays we have developed (oligo-gene array and chromosome 4 array) will provide valuable genomic resources for studying the effects of polyploidy. We are using these microarrays to detect: (1) gene expression in auto- and alloploids; (2) genomic rearrangements; (3) changes in DNA methylation, chromatin structure and transposons; and (4) gene expression in related species.

The analysis of autopolyploid series will allow us to identify gene expression differences that depend largely on changes in genome dosages. For example, in one study microarrays were hybridized using four-dye swaps with cDNA probes prepared from RNAs extracted from leaves of an A. thaliana diploid (2x) and a matched autotetraploid (4x) (Fig. 1 and Table 1). For each dye swap, cDNA from one line was labelled with the fluorescent dye Cy3 (green) and cDNA from the other line was labelled with Cy5 (red) using an indirect incorporation method (Randolph & Waggoner, 1997), and the slide was hybridized with an equal amount of two probes. The cDNA samples also were swapped, each sample was labelled with the other dye and a mixture of the probes was hybridized to another slides. This procedure was repeated for the subsequent dye-swaps using a total of eight slides in each experiment, taking the RNA preparations and printing patterns into account (Table 1). After washing, the slides were scanned (Axon Instruments, CA, USA) and hybridization intensities were recorded. The hybridization intensities reveal whether a gene is equally expressed (yellow), over-expressed (green) or under-expressed (red) in the autotetraploid compared with the diploid (Fig. 3). However, as shown previously (Lee et al., 2004) both dye labelling and slide effects contribute to variation in hybridization intensity. Thus, it is important to replicate these factors and include them in the model used for statistical analysis. This experiment included eight slides; each slide had 26 690 features analysed, of which 26 090 were geneoligos and 600 controls, providing a total of 427 040 data points (26 690 features × 2 RNAs × 4 dye-swaps × 2 printing patterns). It is notable that the data collected from controls are not essential (are often excluded) in the analysis because the controls and data normalization are not needed in the linear model (Kerr & Churchill, 2001; Black & Doerge, 2002; Lee et al., 2003). After the data were processed using
logarithm ratios of green and red hybridization signals, a lowess function (Cleveland, 1979) in the R package (Ihaka & Gentleman, 1996) was used to remove non-linear components (e.g. dye effects). The data were then subjected to ANOVA to estimate the significant changes of gene expression between the two treatments on a per-gene basis. A standard t-test statistic was used for this comparison, based on the normality assumption for the residuals. Both the standard false discovery rate (FDR) (Hochberg & Tamhane, 1987) and Holm’s sequential adjustment (Benjamini & Hochberg, 1995) were used to control multiple testing errors. Holm’s comparison provides a strong control of the familywise error rate (FWER) below level \( \alpha \) with greater power than the standard FDR. The significance level \( \alpha = 0.05 \) was chosen for this study. As a result of this analysis, we detected significant changes in the expression of 416 and 146 genes between autotetraploids and diploid lines using either FDRs or FRDs and Holm’s multiple comparison tests, respectively (Fig. 4).

The comparison of gene expression changes between allopolyploids and their progenitors is more complicated because gene expression may be influenced by ploidy changes and/or by homoeologous genome interactions. One may compare transcript differences between a new allotetraploid (e.g. \( A. suecica \)) and an artificial mix of autotetraploids of the two progenitors (\( A. thaliana \) and \( A. arenosa \)) using dye-swap experiments. This comparison would identify gene expression changes due to genome interactions, because all the materials would be at the same ploidy level. Large changes in gene expression, such as the activation of transposons or one homoeologous gene or the silencing of both homoeologous genes, should be easily detected. However, expression changes involving the silencing of only one of the two homoeologous genes may be too small to be detected using microarray experiments. These different expression patterns can subsequently be verified using additional methods, such as quantitative RT-PCR (Lee et al., 2004), CAPS (Konieczny & Ausubel, 1993; Lee & Chen, 2001) and SSCP (Adams et al., 2003) analyses.

Microarray analysis of genomic DNA may allow detection of genomic rearrangements (e.g. deletions and/or insertions) and/or aneuploids (Hughes et al., 2000) during polyploidization. In this case, genomic DNAs from two lines (e.g. a diploid and a tetraploid) are used as probes in microarray hybridization. The ability to detect deletions and insertions in the genome will be dependent on the size of the rearrangement and the fold-change in copy number. For example, the loss of only one of two homoeologous sequences may not be detectable by microarray analyses. As with expression analyses, replication and the use of statistical models will be important for detecting small fold-changes.

Microarrays will be useful for detecting epigenetic changes involving DNA methylation and chromatin structure, and their association with changes in gene expression and polyploidy. To determine a causal relationship between changes in transcription and chromatin structure, the microarrays are hybridized with genomic DNA fragments of polyploids selected either by chromatin immunoprecipitation (Iyer et al., 2001; Gendrel et al., 2002) or by digestion with methylation-sensitive enzymes (van Steensel, Delrow & Henikoff, 2001). The results from these experiments can be compared with transcription profiling results of the same polyploids to identify relationships between epigenetic and gene expression changes. The chromosome 4 microarray offers some advantages for this study because it includes every genomic fragment in a chromosome and it may detect expression of genes that are not annotated in the databases. The information from oligo-gene microarray analysis is limited to the 3’ end of the genes, but it covers the entire genome. Additional studies could be conducted to elucidate further the likely players in the epigenetic control of polyploid gene expression. These studies might involve comparison of changes in polyploids to those occurring in mutants of chromatin modification or the RNA interference pathway.
Finally, the microarrays developed using Arabidopsis sequences are applicable to gene expression studies in related species, because the sequence identities between A. thaliana and A. arenosa are >92% (Hanfstingl et al., 1994; Henikoff & Comai, 1998; Lee & Chen, 2001) and between A. thaliana and Brassica are >85% (Cavell et al., 1998). Indeed, in a pilot microarray experiment, >95% of Arabidopsis oligos were detected by Brassica probes; >90% of hybridization intensities in each feature were detected when the Arabidopsis microarrays were hybridized with Brassica cDNAs (Lee et al., 2004). Moreover, the sources of variation detected using Arabidopsis and Brassica were very similar, suggesting a common utility of Arabidopsis microarrays for gene expression studies in related species. It is conceivable that gene expression studies can be widely applied to study the population dynamics and speciation mechanisms of polyploids, as demonstrated for species related to yeast and Drosophila (Hartl et al., 2003; Ranz et al., 2003; Rifkin, Kim & White, 2003; Townsend, Cavalieri & Hartl, 2003).

STUDYING MECHANISMS OF DIFFERENTIAL GENE EXPRESSION IN POLYPLOIDS USING ARABIDOPSIS AS A MODEL SYSTEM

It is difficult to study mechanistic aspects of polyploidy effects on animals because of the lack of genetic systems and materials. However, the Arabidopsis polyploid model system provides unique and powerful genetic and genomic resources for elucidating mechanisms of gene and genome duplication. Genome-wide changes in gene expression can be compared with the changes in genome organization and chromatin structure in auto- and allopolyploids, which will provide a global and mechanistic view of polyploidy effects on the regulation of homoeologous genes. Molecular mechanisms of differential gene expression in polyploids may be analogous to X-chromosome inactivation (Lee & Jaenisch, 1997; Avner & Heard, 2001) and imprinting (Martienssen, 1998; Tilghman, 1999; Kelley & Kuroda, 2000; Grossniklaus et al., 2001; Bushell et al., 2003). Indeed, it was shown that parental genomic imbalance and DNA methylation are involved in endosperm development and establishing post-zygotic hybridization barriers between A. thaliana and A. arenosa (Bushell et al., 2003). Moreover, DNA methylation is affected in the interspecific hybrids of marsupials (O’Neill, O’Neill & Graves, 1998), whereas some imprinting genes are misregulated in the interspecific hybrids of North American rodents (Vrana et al., 2000). However, regulation of homoeologous genes in polyploids has some unique features. First, in contrast to X-chromosome inactivation, differential gene expression in polyploids does not occur at every locus on a chromosome or even in small chromosomal segments (Lee & Chen, 2001; Lewis & Pikaard, 2001). Second, gene activation and silencing may result from locus interactions, as compared with allelic interactions imprinting (Tilghman, 1999; Kelley & Kuroda, 2000) and paramutation (Brink, 1973; Hollick, Dorweiler & Chandler, 1997; Herman et al., 2003), although paramutation-like changes of transgene expression have been observed in Arabidopsis autotetraploids (Mittelsten Scheid, Afsar & Paszkowski, 2003). Third, the differentially expressed genes are distributed among all chromosomes (Lee & Chen, 2001; Wang et al., 2004), suggesting that many potential cis- and trans-acting factors and other mechanisms, such as RNA interference, are involved in the initiation and maintenance of silencing. Indeed, RNA silencing was observed with increasing dosage of a transgene in Drosophila (Pal-Bhadra, Bhadra & Birchler, 2002) and transgene activation was dependent on ploidy levels in Arabidopsis (Mittelsten Scheid et al., 1996). Similar phenomena may occur for some homoeologous or duplicate genes in polyploids.

Although previous research has provided evidence that several mechanisms can cause novel gene expression in polyploids (Osborn et al., 2003), many questions remain unanswered regarding the functioning of these mechanisms and their importance in polyploid evolution. The development of a model system involving Arabidopsis polyploids and genomic tools will undoubtedly be very useful in answering some of these questions. The Arabidopsis model system will illuminate our understanding of mechanisms and evolutionary consequences of
polyploidization that will be applicable to other polyploid taxa and natural populations, as well as provide insights into manipulating the expression of duplicate genes in polyploid agricultural crops. Elucidating mechanisms for the function of duplicate genes may ultimately reveal new ways of reactivating or silencing endogenous genes without a need for ‘alien genes’, which may lead the way to improve public acceptance and future applications of biotechnology in agriculture and medicine.

Acknowledgements

We thank James A. Birchler and colleagues of the polyploid consortium project (http://polyploid.agronomy.wisc.edu/) for helpful comments and the National Science Foundation Plant Genome Program (0077774) for support of polyploid research in our laboratories. Work in the Chen laboratory (http://polyploidy.tamu.edu) was supported in part by a National Institutes of Health grant (GM067015) and the Texas Agricultural Experiment Station. The opinions expressed are those of the authors and do not reflect the official policy of the National Institutes of Health, National Science Foundation or the US and state government.

References


Black, MA. PhD thesis. Purdue University; 2002. Statistical issues in the design and analysis of spotted microarray experiments.


Figure 1.
Production of matched autotetraploids, triploids and diploids of Arabidopsis thaliana. Schematic drawing of colchicine induction of polyploids in the G₀, G₁ and G₂ generations. Note that the production efficiency of autotetraploids by colchicine is not 100%. Therefore, tetraploids and diploids can be identified from the same inbred line in the G₂ generation and maintained. Triploids can be generated by crossing a diploid with an autotetraploid. Each line in a ploidy class is derived from a different parent in G₁.
Figure 2.
Arabidopsis allotetraploids. A, schematic karyotypes of parents and F₁ hybrid lines. The centromeres of five Arabidopsis thaliana chromosomes are shown in red, whereas the centromeres of eight A. arenosa chromosomes are shown in green. Independent lineages of allotetraploids are produced by selfing multiple F₁ hybrid lines (indicated by straight arrows). Allopolyploids self-pollinate spontaneously (or can be manually self-pollinated) and progressive inbreeding is expected in advanced generations (S₂ to S₅ and beyond). B, phenotypic variation was observed in seedlings (2 weeks old) of the newly formed allotetraploid lines (S₅ generation) derived from three independent lineages. C, the presence of A. thaliana (At) and A. arenosa (Aa) centromeres in meiotic cells (metaphase I) of a synthetic.
allotetraploid line (left). The *in situ* hybridization was performed using At and Aa centromeric repeats as probes (Comai, Tyagi & Lysak, 2003) (shown in red and green, respectively). The same nucleus stained by 4',6'-diamidino-2-phenylindole hydrochloride (DAPI) is also shown.
Figure 3.
Arabidopsis 26 k oligo-gene microarray. The slide contains a total of 27,648 features printed in 48 subarrays ($576 \times 48 = 27,648$), including 26,090 70-mer Arabidopsis oligos plus controls. The slide was hybridized to cDNA probes prepared from leaves of diploid and isogenic autotetraploid plants and labelled with Cy3 and Cy5, respectively. A subarray of 576 features showing hybridization patterns of genes that are expressed equally (yellow and white), higher in the diploid (green) or higher in the autotetraploid (red).
Figure 4.
Scatter plots showing mean log-fold changes detected using oligo-gene microarrays (27,648 features per slide) probed with Cy3- and Cy5-labeled cDNAs prepared from a pair of Arabidopsis thaliana diploid and autotetraploid lines. The data from eight slide replications (Table 1 and a total of 427,040 data points, see text) were analysed using one-way ANOVA (Black & Doerge, 2002; Black, 2002). The significant genes detected are shown in ‘pluses’ and circles using multiple comparison tests based on the false discovery rate (FDR) (Benjamini & Hochberg, 1995) or FDR and Holm’s sequential adjustment (Hochberg & Tamhane, 1987), respectively, at $\alpha = 0.05$. A detailed description of the statistical model and analysis is given in the text and elsewhere (Lee et al., 2004).
<table>
<thead>
<tr>
<th>Slide no.</th>
<th>RNA or cDNA</th>
<th>Cy3</th>
<th>Cy5</th>
<th>Printing pattern</th>
<th>Dye-swap</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RNA1</td>
<td>Diploid (2x)</td>
<td>autotetraploid (4x)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>RNA1</td>
<td>autotetraploid (4x)</td>
<td>diploid (2x)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>RNA1</td>
<td>diploid (2x)</td>
<td>autotetraploid (4x)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>RNA1</td>
<td>autotetraploid (4x)</td>
<td>diploid (2x)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>RNA2</td>
<td>diploid (2x)</td>
<td>autotetraploid (4x)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>RNA2</td>
<td>autotetraploid (4x)</td>
<td>diploid (2x)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>RNA2</td>
<td>diploid (2x)</td>
<td>autotetraploid (4x)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>RNA2</td>
<td>autotetraploid (4x)</td>
<td>diploid (2x)</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>