There are several reasons why qualitative and quantitative determination of transcript patterns in plant cells is of importance to plant molecular biology. By comparing the concentrations of individual mRNAs present in samples originating from different genotypes, developmental stages or growth conditions, genes can be identified that are differentially expressed and hence may have specific metabolic or morphogenetic functions. As sequence information grows very rapidly, so does the number of protein coding sequences with unknown functions. When the yeast genome project was completed, only approx. 30% of its approx. 6000 coding sequences corresponded to known yeast proteins. More than 30% were orphans sharing no sequence homology to any known gene (Dujon, 1996). At present the function of about half of the yeast genome’s open reading frames is still not known. Analysis of transcript patterns should be valuable in assessing roles of novel sequences in an organism, since the similarity of expression patterns of genes with a known function may indicate functional homology. For example, the first large scale expression profiles of light- and dark-grown Arabidopsis thaliana seedlings obtained by membrane and microarray hybridization revealed numerous genes that were highly regulated by light, but did not have a match in the nucleotide sequence databases. The fact that they are light-inducible might indicate that these genes have as yet undefined photomorphogenetic functions. Isolation of plant genes expressed under specific conditions allows their promoters to be sequenced and compared. Common cis-elements may be localized and their presence may be correlated with specific features of the expression profile of the corresponding genes. This allows analysis of how promoter activity is determined by hierarchical and combinatorial interactions of the signal elements embedded within the promoter sequence. By comparing transcript patterns with proteome data, it may be possible to determine whether the intracellular concentration of specific proteins is preferentially regulated at the level of transcription or by post-transcriptional mechanisms such as mRNA translation efficiency or partitioning between subcellular compartments.

For many years the isolation of genes for which products and mutants were not known was only possible by differential screening of cDNA libraries. The first in vitro technique for the determination of transcript patterns was differential display reverse transcription PCR (DDRT-PCR) developed by Liang and Pardee (1992). For the first time it was possible to determine simultaneously a large part of the transcripts present in a eukaryotic cell within a single experiment with high sensitivity. The technique was applied widely, and for several years no other method was available by which comprehensive transcript patterns of eukaryotic cells could be obtained. Fischer et al. (1995) combined DDRT-PCR and amplified fragment length polymorphism (AFLP), a method developed by Vos et al. (1995) for the characterization of genomic DNA. The new technique, termed restriction fragment length polymorphism-coupled domain directed differential display (RC4D), provided a

Received: 28 July 2000 Returned for revision: 22 September 2000 Accepted: 10 October 2000 Published electronically: 18 December 2000
useful tool to detect differentially expressed members of individual gene families. The cDNA-AFLP technique (Bachem et al., 1996) is based on the selective PCR-amplification of adapter-ligated restriction fragments derived from cDNA. In SAGE (serial analysis of gene expression), developed by Velulescu et al. (1995), short fragments termed tags are extracted from cDNA molecules, concatenated, cloned and sequenced. The abundance of a tag in the sequence reflects the abundance of the corresponding mRNA in the tissue from which the cDNA was prepared. The era of microarray hybridization technology began with the simultaneous quantitative determination of mRNA concentrations of a small set of Arabidopsis genes by a cDNA microarray (Schena et al., 1995). DNA chip technology has now developed to a stage where the state of expression of complete genomes can be recorded with high accuracy on a single chip. Yet this technique did not foreclose the development of other quantitative large scale screening methods. The throughput of the most advanced AFLP and SAGE protocols is comparable to microarrays.

This review summarizes new and old display techniques available to determine transcript patterns and to identify differentially expressed genes in plants. The molecular principles of each method are described, and the biological significance of differential display data obtained from plant systems is discussed using a few typical examples. Performance criteria such as reproducibility, accuracy and throughput are compared to highlight the power of each approach and to delimit its range of application.

MEMBRANE ARRAY HYBRIDIZATION ANALYSIS

Principle

cDNA membrane arrays (macroarrays) are nylon membranes on which denatured double-stranded (ds) DNA fragments are deposited at indexed locations with high speed arraying machines. With the technology currently available it is feasible to array more than 6000 elements on a 12 × 18 cm nylon membrane. The basic principle of acquiring gene expression data using membrane arrays consists of hybridization of an unknown sample to the ordered set of immobilized molecules on the array surface. This produces a specific hybridization pattern that can be read and compared to the pattern produced by a different sample or to a given standard. Probes are prepared by reverse transcription of Poly(A)+ RNA in the presence of 33P-dATP to measure transcript concentrations by hybridization to a macroarray. The membrane is then subjected to a standard hybridization procedure. After washing off the unbound probe the array is read by an imager. Identification and quantitation of the hybridization signals, as well as the subtraction of local background values, is done by image processing software. The use of radioactive isotopes can be circumvented by employing a non-radioactive single or dual colour detection system in which gene expression levels are represented by colour intensities generated by colourimetric reactions (Chen et al., 1998).

Applications

The usefulness of cDNA macroarrays for monitoring gene expression in higher plants has been demonstrated by Desprez et al. (1998), who used a collection of 432 partially sequenced cDNAs arrayed on a nylon support. The filters were employed in hybridization studies with 33P labelled complex probes prepared from mRNA of light- and dark-grown Arabidopsis thaliana seedlings. Only a very small number of mRNAs, corresponding to 1 % of the cDNAs tested, was found to be represented at levels higher than 0.5 % of the total mRNA level. Comparison of light- and dark-grown seedlings showed significant differences in the expression level for approx. 15 % of the genes investigated. For a number of genes previously reported to be regulated by light, differential expression in dark- and light-grown seedlings was observed. These included the chlorophyll a/b binding proteins, plastocyanin and ribulose-1,5-bisphosphate carboxylase. Quantitation of the hybridization pattern also revealed massive variations in expression of many genes for which no differential expression between light- and dark-grown seedlings had previously been observed. Large scale gene expression studies using high density membranes have been reported not only for plants but also for other organisms (Bernard et al., 1996; Pietu et al., 1996).

cDNA MICROARRAY HYBRIDIZATION ANALYSIS

Principle

The array elements of cDNA microarrays are single-stranded DNAs. These arrays are produced by robotically depositing denatured target DNAs, e.g. PCR-amplified inserts from cDNA or EST clones, at defined locations onto microscope slides. The current density of robotic printing allows the design of arrays containing 20 000 cDNA targets (Schena et al., 1995; Kehoe et al., 1999). This density is much higher than in the case of nylon filters, which in theory makes it possible to display most or all transcripts present in a plant cell on a standard microscope slide. The basic steps of preparing and reading cDNA microarrays are depicted in Fig. 1.

Applications

cDNA microarrays were first developed for Arabidopsis thaliana (Schena et al., 1995). Differential gene expression was measured in roots and shoots using a microarray of 48 duplicate cDNA elements simultaneously assayed with a mixture of two probes labelled with fluorescein and lissamine. Comparison of scans obtained with leaf and root mRNA revealed 26 genes differing more than five-fold in expression between root and leaf tissue. mRNA from the light regulated CABI gene was approx. 500-fold more abundant in leaf than in root tissue. In a transgenic line overexpressing the HAT4 transcription factor, the HAT4 gene was expressed at a level 60-times that of the wild type. Most of the other genes tested gave expression levels within a factor of five between the two samples.
Ruan et al. (1998) analysed cDNA microarrays containing 1443 Arabidopsis thaliana genes for expression profiles in different organs and at different developmental stages. More than 80% of the DNA elements on the arrays generated detectable fluorescence signals. Among the detectable elements only 1.4–5% represented highly expressed genes with abundance of more than 100–500 transcripts per cell. In this class, as expected, many members were well characterized housekeeping or tissue-specific genes. The percentage of genes expressed at modest levels (i.e. between ten–50 transcripts per cell) was between 11–16%, depending on the organ tested. The majority of the expressed genes were low abundance with levels of less than ten–50 transcripts per cell. Hence the transcript complexity in different tissues of Arabidopsis was found to be consistent with early studies in other plants using RNA/DNA reassociation kinetics (Kamalay and Goldberg, 1980).

Comparing leaf and root, 34% of the transcripts were expressed at significantly different levels. The most differentially expressed sequences were those known to be involved in photosynthesis. In flowers, many genes were found to be downregulated when compared with their expression patterns in leaves. Conversely, over 200 genes were upregulated in flowers when compared with the leaf. Of the 1443 coding sequences used, 173 had no significant homology to any known cDNA sequences in the current public databases. Two of these exhibited an expression pattern comprising higher transcript abundance in roots and developing flower buds, but lower abundance in mature leaves and opened flowers. This expression pattern was similar to that of a known gene that encodes a G-protein involved in cell division. The proteins encoded by the two novel sequences may thus have functional roles related to cell division in meristematic tissues. This illustrates the
point that functions of as yet uncharacterized genes may be inferred from the similarity of their expression patterns to the patterns of genes of known function.

The state of DNA microarray technology and the technique’s potential have been described in several recent reviews (Hoheisel, 1997; Lemieux et al., 1998; Ramsay, 1998; Kehoe et al., 1999; Khan et al., 1999). Detailed descriptions are also available dealing with technical aspects of producing and reading microarrays (Bowtell, 1999; Cheung et al., 1999; Eisen and Brown, 1999) and with the specialized database software required for storage and retrieval of complex gene expression data (Ermolaeva et al., 1998).

**OLIGONUCLEOTIDE MICROARRAY HYBRIDIZATION ANALYSIS**

**Principle**

The basic principle of using oligonucleotide microarrays was first proposed in the late 1980s, when several groups independently developed the concept of sequencing by hybridization, i.e. determining a DNA sequence by hybridization to a comprehensive set of oligonucleotides, such as all the possible 65 536 octamers (Chetverin and Kramer, 1994). The initial aim was the establishment of a faster, less costly and more effective sequencing approach. When it gradually became apparent that microarrays could also be used for large scale comparative measurements of mRNA levels of sets of genes in different cell types or under different conditions, the interest shifted to transcriptional profiling, which is now the most prominent application.

By automatically printing presynthesized oligonucleotides onto microscope slides, oligonucleotide microarrays can be produced in the same way as cDNA microarrays at densities of up to 30 000 elements cm$^{-2}$ (Yershov et al., 1996). A different way of producing oligonucleotide microarrays is semiconductor photolithography in combination with modified phosphoramidite-based oligonucleotide synthesis chemistry (Fodor et al., 1991; Pease et al., 1994). This technique generates high density microarrays using comparatively few synthesis steps, since the number of probes increases exponentially with a linear increase in the number of synthesis cycles. For example, a complete set of octanucleotides consisting of 65 536 probes can be produced in 32 chemical steps in 8 h (Ramsay, 1998). Photolithography has been used to produce arrays at densities of up to $10^8$ probes per cm$^2$. This corresponds to an element size of 5–10 μm, and it is expected that optimization will permit features of approx. 1 μm. The basic steps of probe labelling, hybridization, signal detection and data processing are the same regardless of array type.

**Applications**

Oligonucleotide microarrays have been used to measure expression levels of genes in mice (Lockhart et al., 1996), yeast (Wodicka et al., 1997) and bacteria (DeSaizieu et al., 1998). So far, oligonucleotide microarray analysis has not yet been reported for a plant. Reasons for this will be discussed later.

**SERIAL ANALYSIS OF GENE EXPRESSION**

**Principle**

Serial analysis of gene expression (SAGE) is another large scale method allowing quantitative and simultaneous analysis of many transcripts (Velculescu et al., 1995). From each species present in a cDNA population, a unique sequence tag (approx. 9 bp) is extracted by a series of enzymatic restriction and ligation steps. The cleavage reactions are catalysed by a pair of restriction enzymes, a tetracutter termed the anchoring enzyme and a type II S restriction endonuclease, which is termed the tagging enzyme. The tags are linked together, concatenates are inserted into a plasmid vector, cloned and sequenced. In the sequence, tags are easily distinguished, because they are separated by the recognition sequence of the anchoring enzyme which establishes the register and boundaries of each tag. Hence concatenation of short sequence tags allows the efficient analysis of transcripts in a serial manner by the sequencing of multiple tags within a single clone. A transcription profile is obtained by determining the total frequency of each tag in a SAGE clone library, which directly reflects the abundance of the corresponding mRNA in the tissue. Individual steps of SAGE library construction are shown in Fig. 2.

**Applications**

Serial analysis of gene expression was initially applied to analyse the expression of approx. 1000 pancreatic genes (Velculescu et al., 1995). Since then the power of SAGE for use in gene expression studies has been demonstrated in a number of cases. The first SAGE analysis of a higher plant was completed by Matsumura et al. (1999). More than 10 000 tags derived from approx. 6000 genes expressed in rice were studied. SAGE showed that most of the highly expressed genes in rice seedlings are housekeeping genes. A remarkable exception is a metallothionein gene, whose expression counts for more than 1% of total gene expression. Analysis of approx. 2000 tags obtained from anaerobically treated and untreated seedlings revealed that most genes were expressed at similar levels. Interestingly, the genes required for anaerobic metabolism were not identified in the differentially expressed genes. Using the tag sequence as a primer, a 0.2 kbp cDNA fragment belonging to an elongation factor EF-1α gene could be successfully amplified, thus demonstrating that the sequence of a 9 bp tag plus the adjacent 4 bp tetracutter recognition site can serve as a PCR primer for the recovery of longer cDNA fragments that can in turn be used for the isolation of the corresponding gene.

**THE cDNA-AFLP TECHNIQUE**

**Principle**

cDNA-AFLP with two restriction enzymes. cDNA-AFLP is an RNA fingerprinting technique that evolved from AFLP (amplified fragment length polymorphism), a method described by Vos et al. (1995) for the fingerprinting of genomic DNA. The classical cDNA-AFLP procedure
(Bachem et al., 1996) uses the standard AFLP protocol on a cDNA template. The technique involves three steps: (1) restriction of cDNA and ligation of oligonucleotide adapters; (2) selective amplification of sets of restriction fragments using PCR primers bearing selective nucleotides at the 3' end; and (3) gel analysis of the amplified fragments (Fig. 3). Restriction of plant cDNA with a combination of two restriction enzymes, a tetracutter and a hexacutter, allows a significant fraction of the cDNA population to be cleaved and to be represented as a discrete banding pattern on a sequencing gel. In genomic AFLP with plant DNA, three selective bases on the end of each primer are required to give a useful banding pattern. The lower complexity of cDNA allows the use of two selective bases for each primer giving a total of 256 possible primer combinations. The largest cDNA-AFLP products visible on a polyacrylamide gel are divided in half, ligate to linkers, and then amplify, cleave with NlaIII, isolate ditags, concatenate and clone.
sequencing gel are around 1000 bp in size, the lower end of the gel representing approx. 100 bp. In this size window an average of 40 bands can be observed for each primer combination, corresponding to a total of approx. 10 000 bands.

**cDNA-AFLP with one restriction enzyme.** A systematic comparison of known potato cDNA sequences showed that approx. 45% are cleaved by the *Ase*I/*Taq*I restriction enzyme combination. Thus, in so far as only one pair of

**Fig. 3.** The classical cDNA-AFLP protocol. cDNA is synthesized from total RNA or poly(A)+-RNA and is digested with *Taq*I and *Ase*I, which recognize four and six bp, respectively. A complete digest of plant cDNA with these enzymes produces five different types of molecules: *Ase*I/*Ase*I fragments, *Ase*I/*Taq*I fragments, *Taq*I/*Taq*I fragments and two terminal fragments with only one cohesive end. *Taq*I, which cuts DNA frequently, generates small cDNA fragments (around 256 bp on average), which amplify well and lie in the optimal size range for separation on sequencing gels. *Ase*I, which cuts only rarely due to its longer recognition sequence, reduces the number of fragments to a manageable size. Following digestion, double-stranded adapters are ligated to the restriction fragments to generate templates for amplification. PCR amplification is carried out in two steps. In the first step, around 15 cycles of non-specific amplification are carried out using primers without extensions. The products of this reaction are then subjected to a second round of PCR amplification using primers bearing at their 3' end two additional nucleotides which extend into the sequence of the restriction fragments, allowing only a subpopulation to be amplified. All the 256 possible primer combinations are necessary to amplify the whole cDNA population. The amplicons are separated on a polyacrylamide gel and visualized by autoradiography. Most of the bands represent *Ase*I/*Taq*I fragments because *Ase*I/*Ase*I fragments are rare and *Taq*I/*Taq*I fragments are not visible on the gel. RNA probes from different sources (A, B) will produce different cDNA-AFLP banding patterns, which allows differentially expressed cDNAs to be identified (arrows).
enzymes is applied, about half of the transcripts present in a cell will not be detected by the standard cDNA-AFLP technique. To obtain more comprehensive patterns, Habu et al. (1997) modified the cDNA-AFLP protocol and showed that the rarely cutting enzyme can be omitted, and meaningful banding patterns can be produced using _TaqI_ alone. Samples derived from buds of red and white flowers of the common morning glory (_Ipomoea purpurea_) were compared using 96 different primer combinations, each of which gave approx. 50 bands, corresponding to a total of approx. 5000 bands.

**iAFLP.** _iAFLP_ (introduced AFLP) is a quantitative high throughput expression profiling method specifically designed to measure the concentrations of known transcripts in numerous different probes (Kawamoto et al., 1999). cDNA from each probe is restricted with _MboI_ and ligated to one of up to six adapters having short insertions of various lengths into a common sequence (polymorphic adapters). Following ligation, the differentially adapted cDNAs are pooled and 3' end fragments are selectively amplified with a gene specific primer and a fluorescently labelled adapter primer. The amplicon is then separated on an automatic sequencer. Due to length heterogeneity introduced by the polymorphic adapters, _iAFLP_ fragments from different probes will produce distinct peaks on the electropherogram. Transcript abundance is determined by evaluating peak areas relative to an internal standard.

**Applications**

cDNA-AFLP and its application to plants was first described by Bachem et al. (1996), who analysed differential gene expression in a synchronized potato _in vitro_ tuberization system. During screening with different primer combinations, two lipoxgenase cDNA fragments were isolated on the basis of their differential expression during potato tuber formation. Both transcripts are highly tuber-specific and are expressed strongly in 15-d-old tubers, but not in stolons, leaves or petioles and only at very low levels in stems. The dramatic induction of a lipoxgenase gene just after the start of tuberization led the authors to speculate that the expression of at least one of these enzymes might directly be linked to the tuber development process.

Following this initial report, a small number of papers have described the use of cDNA-AFLP fingerprinting in plant and animal systems. Habu et al. (1997) compared mRNA samples obtained from the flower buds of two lines of _Ipomoea purpurea_. Fourteen cDNA fragments (approx. 0.3 %) amplified differently in the two samples. Two of these were shown to have been derived from a gene that was actively expressed in the buds of red flowers but not in those of white flowers. Sequence analysis showed that this cDNA carries a sequence highly homologous to the chalcone synthase gene, a key enzyme in the flavonoid biosynthetic pathway. cDNA-AFLP was also applied to identify differentially expressed genes in cold-tolerant and cold-sensitive alfalfa genotypes and rice (Ivashuta et al., 1999).

**DIFFERENTIAL DISPLAY REVERSE TRANSCRIPTION PCR**

**Principle**

DDRT-PCR, the differential display reverse transcription PCR method, was first described by Liang and Pardee (1992). The method was intended to provide an effective tool to detect individual mRNA species that are differentially expressed in different eukaryotic cell types and then to permit recovery and cloning of the corresponding cDNAs. The principle of the DDRT-PCR technique is to amplify specific subpopulations of mRNA, using reverse transcriptase and PCR to produce a population of PCR fragments of different lengths. To do this, mRNA is reverse-transcribed in subsets using 12 anchored oligo(dT) primers that recognize different fractions of the total mRNA population. The resulting cDNA is then amplified with the same anchored oligo(dT) primer and a short arbitrary primer. Labelled dATP is introduced in the reaction and the labelled products are separated on a DNA sequencing gel and visualized by autoradiography.

Many refinements of the original differential display technology have been described. Instead of decameric upstream primers, elongated arbitrary primers have been used. This opens the possibility of including common sequence motifs to target particular classes or families of genes (Donohue et al., 1995; Johnson et al., 1996; Gonsky et al., 1997). Methods employed for the separation of DDRT-PCR products include denaturing and non-denaturing polyacrylamide gel electrophoresis (Liang et al., 1995), capillary electrophoresis (George et al., 1997) and even electrophoresis on standard agarose gels (Rompf and Kahl, 1997). To label DDRT-PCR products, [³²⁵P]dATP is generally used (Simon and Oppenheimer, 1996). Alternatives to radioactive labelling are fluorescence detection (Bauer et al., 1993; Jones et al., 1997), silver staining (Simon and Oppenheimer, 1996; Gottschlich et al., 1997) and chemoluminescence detection (An et al., 1996). To circumvent subcloning, direct sequencing approaches for differential display products have been developed (Reeves et al., 1995; Wang and Feuerstein, 1995; Simon and Oppenheimer, 1996; Buess et al., 1997).

**Applications**

Due to its technical simplicity the DDRT-PCR method has become very popular. Since its introduction, hundreds of applications ranging from microorganisms to humans have been described. I will consider the identification of an Arabidopsis gene (NAP) which is an immediate target of the floral homoeotic genes _APETALA_ and _PISTILLATA_ (Sablowski and Meyerowitz, 1998). Homoeotic genes encode transcription factors, which are mostly MADS box proteins in plants. In _Arabidopsis_ flowers both _APETALA_ (_AP3_) and _PISTILLATA_ (_PI_)) are expressed throughout the development of petals and stamens. The _AP3_ and _PI_ proteins bind DNA _in vitro_ only as an _AP3/PI_ heterodimer, which probably acts as a master regulator necessary and sufficient to set in motion the programs of gene expression required for petal and stamen
development. To identify AP3/PI target genes, transgenic Arabidopsis plants were constructed exhibiting constitutive expression of PI combined with ubiquitous expression of a steroid regulated version of AP3. In this artificial differential expression system the mechanism of AP3 activation requires no protein synthesis. Thus, the immediate effects of AP3/PI activation can be analysed by differential display while all downstream gene regulatory events are blocked by a protein synthesis inhibitor. To identify mRNAs affected by activation of the AP3 product, mutant flowers were treated with dexamethasone and mRNA populations were compared by DDRT-PCR. For one candidate (NAP), RNA blot analysis showed that the corresponding 1–3 kb mRNA was more abundant in wild type than in mutant flowers, suggesting activation by AP3. The expression pattern of NAP and the phenotypes caused by its mis-expression suggest that it functions in the transition between growth by cell division and cell expansion in stamens and petals. In Arabidopsis, DDRT-PCR thus served to detect an important genetic link between homeoectic gene activity and flower morphogenesis.

Further examples of the successful isolation of cDNAs associated with distinct stages of plant growth and development include genes expressed during the plant cell cycle (Callard and Mazzolini, 1997), embryogenesis (Heck et al., 1995), fruit ripening (Tieman and Handa, 1996), tuber formation (Monte et al., 1999) and the development of seeds (Nuccio and Thomas, 1999). The DDRT-PCR technique has also been applied to identify several circadian clock regulated genes in Arabidopsis (Kreps et al., 2000) and to isolate several cDNAs induced in plants under various stress conditions e.g. salt (Muramoto et al., 1999), heat shock (Visioli et al., 1997), ozone treatment (Kiskinen et al., 1997), nutrient starvation (Petrucco et al., 1996), wounding (Titarenko et al., 1997) and pathogen attack (Truesdell and Dickman, 1997).

RFLP-COUPLED DIFFERENTIAL DISPLAY

Principle

Many genes and their protein products have a modular structure where the presence of certain domains (family specific domains, FSDs) defines membership in different gene families. This is well characterized for the chlorophyll a/b binding proteins and for many transcription factors. Restriction fragment length polymorphism-coupled domain directed differential display (RC4D) is a method specifically designed to analyse expression of multigene families at different developmental stages, in diverse tissues or in different organisms (Fischer et al., 1995). RC4D combines cDNA-AFLP technology with a gene family specific version of DDRT-PCR. In RC4D, instead of arbitrary decameric primers, longer primers directed against an FSD are used, allowing cDNAs belonging to the same gene family to be selectively amplified. As the amplification products are relatively uniform in length, restriction fragment length polymorphism (RFLP) is introduced by digestion with a frequently cutting restriction enzyme. This reduces the amplicon size from approx. 1 kbp to several hundred base pairs, which is optimal for separation on acrylamide gels. Family members can thus easily be distinguished by size. The basic steps of RC4D are depicted in Fig. 4.

Applications

RC4D was first used to analyse differential expression of MADS box genes in male and female inflorescences of maize (Fischer et al., 1995; Huang et al., 1996). The name MADS was constructed from the initials of the first four members of the gene family, which were MCM1 (yeast), AGAMOUS (plants), DEFICIENS (plants) and SRF (human). A small collection of MADS box primers was designed, directed against sequences encoding derivatives of a highly conserved amino acid motif which covered all its variations known from plants. RC4D yielded many fragments significantly different in size. Most of them were equally present in both sexes. Four already known and two new MADS box genes were identified, being either specifically expressed in the female sex or preferentially expressed in male or female inflorescences, respectively. The two new MADS box genes belong to a subfamily showing sequence similarity to floral homeoectic and transcription factor genes (Huang et al., 1996).

Another example of using RC4D was provided by Tahtiharju et al. (1997), who identified several cDNAs coding for calcium dependent protein kinases involved in calcium signalling during cold induction of the kin genes of Arabidopsis thaliana.

DIFFERENTIAL SCREENING OF cDNA LIBRARIES

Principle

The oldest method of isolating differentially expressed cDNAs is the differential screening of plasmid or phage cDNA libraries. In principle, differential screening can be used to identify any differentially regulated gene. In practice, however, the method works well only when the mRNA of interest comprises more than 0.05 % of the total mRNA in one cell and less than 0.01 % in the other (Sambrook et al., 1989). The sensitivity with which cDNAs corresponding to relatively rare mRNAs are identified is considerably improved if the concentration of specific sequences in a cDNA probe is increased by subtractive hybridization (SH). There are numerous protocols for SH but the principle remains the same. mRNA extracted from one cell type is used as a template to make radiolabelled cDNA which is then hybridized exhaustively with an approx. 20-fold excess of mRNA extracted from a second type of cell in which the gene of interest is not expressed. cDNAs corresponding to mRNAs expressed in both cell types will form DNA:RNA hybrids, which can be separated from single-stranded DNA by chromatography on hydroxyapatite columns or by streptavidin biotin interaction. The resulting subtracted single-stranded probe
is then used for library screening or for the construction of a subtracted cDNA library. SH detects rare mRNAs with prevalences ranging from 1/500 to 1/200 000 with the median being approx. 1/20 000 (Wan et al., 1996).

**Applications**
A classical example of the identification of a low abundance transcript from a plant by differential screening of cDNA libraries was provided by the isolation of the first

---

**Figure 4.** Schematic representation of the RC4D protocol. cDNA is synthesized from mRNA with an oligo(dT) primer bearing a PCR downstream primer binding sequence at its 5' end. PCR is performed with the downstream primer and an upstream primer specific for a family specific domain (FSD). This results in a mixture of truncated family member cDNAs. The amplicon is digested with a frequently cutting restriction enzyme, and double-stranded linkers are ligated to the cohesive ends. PCR with a linker primer and a FSD primer results in a population of family member cDNA fragments of different lengths. To get rid of the unligated fragments, a further round of PCR is performed using the FSD primer and a primer directed against the linker. Amplification products are then used as a template to extend a radiolabelled FSD primer, and extension products are separated on acrylamide gels. Different probes (A, B) will produce different RC4D banding patterns, which allows differentially expressed cDNAs to be identified (arrows); modified from Fischer et al., 1995.
phytochrome cDNA clones from *Avena sativa* (Hershey et al., 1984). This opened the door for determination of the complete amino acid sequence of the phytochrome apo-protein and the isolation and characterization of phytochrome genes from *Avena* and other plant species (reviewed in Colbert, 1988). Differential screening has been widely applied in plants, including the isolation of cDNA sequences corresponding to genes expressed specifically during flower development in tobacco (Goldman et al., 1992) and *Sinapis alba* (Staiger and Apel, 1993) and the characterization of male flower cDNAs from maize (Wright et al., 1994).

**Potential, limitations and application range of current display techniques**

**Quantitative high throughput display systems**

cDNA array hybridization, oligonucleotide array hybridization, SAGE and iAFLP represent quantitative high throughput display techniques allowing the intracellular concentrations of hundreds or thousands of transcripts to be determined simultaneously in a single experiment. The main characteristics of these techniques are compared in Table 1.

**cDNA arrays.** Macroarray technology is relatively cheap. Since the filters can be rehybridized up to ten times, a small set of filters is sufficient to monitor expression of most or all *Arabidopsis* genes in two different tissues or under two different conditions. In comparison to clones arrayed on membrane filters, microarrays are 50-times more sensitive and permit simultaneous analysis of two different probes. A disadvantage of these arrays is that they are not reusable and the equipment needed is more complex. Sequence information is not absolutely required, since cDNA arrays can be made using anonymous cDNA or EST clones. Those elements associated with interesting transcription patterns can be sequenced after the hybridization experiment has been completed. However, the number of clones required to ensure a high probability that a low abundance mRNA will be present at least once in the cDNA library used to produce the array is very large. More than 20 standard size cDNA microarrays are required in order to have a 99% probability of at least one array element hybridizing to a rare mRNA with prevalence ≤0.001% (Table 1). Because of this, the usefulness of cDNA array technology for the detection of rare transcripts corresponding to unknown genes is limited, despite the technique’s high potential for automation.

**Oligonucleotide microarrays.** A distinct advantage of oligonucleotide microarrays over cDNA microarrays is that preparation and storage of many different DNA species can be avoided by on chip oligonucleotide synthesis. To give a chance for every possible gene fragment derived from an unknown genome to form a perfect hybrid with at least one element of an oligonucleotide microarray, the array should display all possible oligonucleotide sequences of the given length. The array density required to design such arrays is several orders of magnitude higher than the density at which oligonucleotide arrays can be produced with current chip technology. Hence, monitoring expression of most or all genes of an organism by oligonucleotide microarrays requires the genome sequence of that organism to be known. For most plant species extensive sequence data are not available. For the purpose of transcriptional profiling in plants this makes cDNA microarrays the array type of choice.

For a given transcript there is a quite accurate relationship between fluorescence intensity and transcript concentration regardless of array type. Exact quantitative comparisons of different transcripts, however, require that for all probe species hybridizing with the array this relationship is the same. This is questionable, since individual probe species have unique secondary structures, melting temperatures and reassociation kinetics (Southern et al., 1999), which makes hybridization of all probe species under optimum conditions impossible.

**Serial analysis of gene expression.** The power of SAGE is comparable to microarray technology. Like microarrays, SAGE can quantify low abundance transcripts and detect small differences in transcript concentration. Two major advantages of SAGE over microarrays are that uncloned cDNA can be analysed and that no special device other than a sequencer is required. Since SAGE clones may contain 20–50 tags, a representative SAGE library is much smaller than the cDNA library required to build an equivalent microarray. A unique feature of SAGE is that the method has digital output, which makes data processing relatively simple. In order to be represented by a SAGE tag, a cDNA must have at least one recognition site for the anchoring enzyme. Since tetracutters cleave DNA frequently, this should be the case for almost all transcripts. The 9 bp tags generated by the original SAGE procedure together with the adjacent tetracutter sequence are also long enough to identify all transcripts almost unequivocally and provide sufficient sequence information for PCR cloning and screening of cDNA libraries (Matsumura et al., 1999). The number of SAGE tags to be sequenced depends on mRNA abundance and the desired accuracy of quantification. In principle the latter is not limited. If it is desired to count an average of ten tags per cDNA species, which would allow a rough quantitation, approx. 150 000 tags (approx. 1-65 Mb) have to be sequenced. This requires sequencing of 7500 clones, each clone containing 20 tags on average (Table 1). Because of several technical difficulties and inherent drawbacks, SAGE has been used less frequently for transcript analysis than DNA microarray hybridization. Meanwhile these problems have largely been alleviated by various modifications and improvements of the original SAGE protocol (Powell, 1998; Datson et al., 1999; Kenzelmann and Muhlemann, 1999; Peters et al., 1999; Ryo et al., 2000).

**Introduced AFLP.** Throughput and accuracy of iAFLP are comparable to microarrays. Three different primer colours can be used, allowing up to 288 PCR reactions to be
<table>
<thead>
<tr>
<th></th>
<th>Macroarray</th>
<th>cDNA microarray</th>
<th>Oligonucleotide microarray</th>
<th>SAGE</th>
<th>iAFLP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principle</strong></td>
<td>Nucleic acid hybridization</td>
<td>Nucleic acid hybridization</td>
<td>Nucleic acid hybridization</td>
<td>Extracting short sequence</td>
<td>Selective amplification of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tags from cDNA; cloning,</td>
<td>adapter ligated 3' cDNA ends</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sequencing and counting tags</td>
<td></td>
</tr>
<tr>
<td><strong>Equipment</strong></td>
<td>Robot, 2D imager, standard image</td>
<td>Advanced robotics, array</td>
<td>Photolithography, array</td>
<td>Sequencer</td>
<td>Sequencer</td>
</tr>
<tr>
<td></td>
<td>processing software</td>
<td>scanner, specialized data</td>
<td>scanner, specialized data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>processing software</td>
<td>processing software</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Density</strong></td>
<td>6000</td>
<td>30 000</td>
<td>250 000</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>Elements/clones to analyse</strong></td>
<td>460 000</td>
<td>460 000</td>
<td>≤ Number of protein coding</td>
<td>7500</td>
<td>=Number of genes × probes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>genes in target genome</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reproducibility (%)</strong></td>
<td>±30</td>
<td>±50</td>
<td>±50</td>
<td>Increases with amount of</td>
<td>±4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sequencing</td>
<td></td>
</tr>
<tr>
<td><strong>Sensitivity (%)</strong></td>
<td>&lt;0.01</td>
<td>0.5 × 10⁻⁶–10⁻⁵</td>
<td>2 × 10⁻⁴–2 × 10⁻³</td>
<td>Increases with amount of</td>
<td>&lt;10⁻⁵</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sequencing</td>
<td></td>
</tr>
<tr>
<td><strong>Total RNA (µg)</strong></td>
<td>25</td>
<td>50–200</td>
<td>25–150</td>
<td>100</td>
<td>25–100</td>
</tr>
<tr>
<td><strong>Considerations</strong></td>
<td>Reusable; many cDNA clones to be</td>
<td>Many cDNA clones to be</td>
<td>Requires extensive sequence</td>
<td>Generates sequence</td>
<td>cDNA or EST Sequence</td>
</tr>
<tr>
<td></td>
<td>maintained</td>
<td>maintained</td>
<td>information about target</td>
<td>information; SAGE library</td>
<td>must be known</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>genome</td>
<td>construction difficult</td>
<td></td>
</tr>
<tr>
<td><strong>Reference</strong></td>
<td>Desprez et al., 1998</td>
<td>Lockhart et al., 1996;</td>
<td>Wodicka et al., 1997;</td>
<td>Velculescu et al., 1995;</td>
<td>Kawamoto et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Schena et al., 1996;</td>
<td>de Saizieu et al., 1998</td>
<td>Datson et al., 1999</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ruan et al., 1998</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The number \((N)\) of random clones necessary to obtain a near to complete array hybridization expression profile from one cell type grown under defined conditions is calculated by \(N = \ln(1 - P)/\ln(1 - 1/n)\) using \(P = 0.99, 1/n = 10^{-5}\); \(P\) = probability that a low abundance mRNA will be present at least once in the library used to produce the array; \(1/n\) = fraction of the total mRNA that is represented by a single type of rare mRNA (Sambrook et al., 1989). The number of SAGE clones required to obtain a rough SAGE profile is estimated as follows: \(N = ZM/k\), where \(Z\) is the average number of tags per mRNA species (10), \(M\) the number of mRNA species per cell (15 000) and \(k\) the number of tags per clone (20). Reproducibility: estimated confidence interval for repeated measurements. Sensitivity is given in terms of mRNA abundance (mass of desired mRNA/mass of total mRNA in RNA probe). Total RNA required is for one experiment. Protocols employing RNA or cDNA amplification may require 500–5000-fold less material.
analysed in a single sequencer run. Thus, if profiles from six different tissues are analysed, the throughput is 1728 (288 × 6) data points per run or 5184 data points per day (three runs). Control measurements performed with defined template mixes indicate that the method reflects the relations between the ratio of polymorphic products and the ratio of targets in the initial mRNA population quite accurately. The main limitation of iAFLP is that only known transcripts can be quantified, since a primer binding sequence is required downstream of the MboI site closest to the cDNA’s 3’ end.

Qualitative display systems

The most characteristic features of cDNA-AFLP, DDRT-PCR, RC4D and differential screening are summarized in Table 2. With the exception of RC4D, which requires the nucleotide sequence of a gene family specific domain, these methods can be used for the identification of any differentially expressed gene. They are, however, less accurate and allow transcription profiles to be determined only on a qualitative or semiquantitative basis.

cDNA-AFLP. In cDNA-AFLP, universal adapters are added to DNA fragments produced by restriction of cDNA with one or two restriction enzymes. The method needs only minute amounts of RNA due to the preamplification step performed with non-selective primers. Because stringent hybridization conditions are used in the amplification reactions, mismatched priming events are observed only in cases where transcript levels are extremely high. This results in cDNA-AFLP banding patterns being highly reproducible and almost free of false positives. cDNA-AFLP provides a straightforward verification mechanism of band identity and homogeneity. When using AFLP primers with additional base extensions it is possible to selectively amplify one cDNA fragment, thereby eliminating virtually all other fragments from being co-amplified. Because band intensity is a direct function of template concentration, it is likely that transcript concentrations can be measured quite accurately by combining cDNA-AFLP with high resolution quantitative separation techniques.

A significant disadvantage of the cDNA-AFLP method is the requirement for appropriate restriction sites on the cDNA molecules. cDNA-AFLP experiments using different enzymes are necessary to visualize every cDNA present in a plant cell with high probability. For several reasons, this is also true if the one enzyme variant of the cDNA-AFLP protocol is used (Habu et al., 1997).

DDRT-PCR. The DDRT-PCR technique has been widely used to isolate differentially expressed plant genes due to its technical simplicity and the ease with which the primary data are obtained. However, the method suffers from severe drawbacks. Even the most advanced procedures generate at least 30% false positives that do not represent differentially expressed genes (Malhotra et al., 1998). This makes a downstream verification process necessary, which is not only labour intensive, but also requires large amounts of RNA that often cannot be obtained without RNA amplification (Poiber et al., 1997). In theory, DDRT-PCR should detect more than 95% of the transcripts present in a eukaryotic cell (Bauer et al., 1993). Practically, however, the sensitivity of the method depends on both primer sequence and concentrations of competing template binding sites (Bertioli et al., 1995; Wan et al., 1996). Thus it seems difficult to predict whether or not a given primer set will detect a specific transcript, even if abundant, in an uncharacterized cDNA population, where concentrations of competitors are not known. Because amplification starts from the 3′ end of the mRNA, amplicons are from the untranslated 3-region of the gene and often do not contain the information required to perform successful similarity searches in sequence databases. Finally, in DDRT-PCR, there is no accurate relationship between signal strength and the initial concentration of the corresponding mRNA, which precludes quantitative measurement of transcript concentration. A few systematic comparisons of cDNA-AFLP and DDRT-PCR have been published, indicating that cDNA-AFLP is superior to DDRT-PCR in several respects (Habu et al., 1997; Jones and Harrower, 1998).

RC4D. The use of longer primers and stringent annealing temperatures makes the RC4D method more reproducible and far less parameter dependent than DDRT-PCR, allowing semiquantitative determination of transcript concentration. The main limitation of the RC4D method is that it requires a conserved domain to be situated upstream of the gene’s 3′ end, which only allows genes belonging to the same gene family to be analysed. Family members bearing a respective restriction site within the FSD will be lost because cleavage of the amplicon resulting from the initial PCR amplification will destroy the domain specific primer binding site.

Differential screening

Because of the many steps required to construct a representative library and the necessity to screen thousands or millions of clones, differential screening is generally considered to be extremely laborious and inefficient. If one is looking for rare mRNAs, this is certainly true. However, in cases where a gene is strongly expressed under a given set of conditions and only weakly expressed under others, the number of clones that have to be screened to find a certain cDNA may be quite small. For example, the probability of an mRNA with prevalence 1% being represented at least once in a cDNA library of 500 clones is >99% (Table 1). The usefulness of subtractive hybridization (SH) is often limited by the inability to obtain sufficient mRNA to drive SH to completion and by ineffective hybridization. As with DDRT-PCR, a significant fraction of the candidate clones isolated by SH consists of false positives. There is also a correlation between clone redundancy and the fraction of false positives, indicating that SH identifies abundant mRNAs more easily than rare mRNAs. SH works in only one direction, i.e. to identify mRNAs expressed only in those cells from which subtractor RNA has been prepared, a second experiment is required (Wan et al., 1996).
**Table 2. Qualitative display systems**

<table>
<thead>
<tr>
<th></th>
<th>cDNA-AFLP</th>
<th>DDRT-PCR</th>
<th>RC4D</th>
<th>DS</th>
<th>SH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principle</strong></td>
<td>Selective amplification of adapter-ligated cDNA</td>
<td>Selective amplification of 3’ cDNA ends</td>
<td>Combines cDNA-AFLP and DDRT-PCR</td>
<td>Nucleic acid hybridization</td>
<td>Nucleic acid hybridization</td>
</tr>
<tr>
<td><strong>Equipment</strong></td>
<td>PAGE or sequencer or CE</td>
<td>PAGE or sequencer or CE</td>
<td>PAGE</td>
<td>Standard</td>
<td>Standard</td>
</tr>
<tr>
<td><strong>Bands or clones to screen</strong></td>
<td>10 000</td>
<td>38 000</td>
<td>Few, depending on size of gene family</td>
<td>460 000</td>
<td>46 000</td>
</tr>
<tr>
<td><strong>Reproducibility (%)</strong></td>
<td>100</td>
<td>42</td>
<td>100</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>Sensitivity (%)</strong></td>
<td>$2 \times 10^{-5}$</td>
<td>$2 \times 10^{-5}$ to $10^{-2}$, depending on primer sequence and genetic background</td>
<td>$2 \times 10^{-5}$</td>
<td>$&gt; 0.05$</td>
<td>$5 \times 10^{-5}$ to $5 \times 10^{-3}$</td>
</tr>
<tr>
<td><strong>Total RNA (µg)</strong></td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>Considerations</strong></td>
<td>Method may generate heterogeneous bands, but provides a simple mechanism of band verification</td>
<td>Generates at least 30% false positives; short 3’ end fragments prevail; biased towards abundant transcripts; bands usually heterogeneous</td>
<td>Requires sequence of a family specific domain</td>
<td>Generates false positives</td>
<td>Generates false positives; technically difficult</td>
</tr>
<tr>
<td><strong>Reference</strong></td>
<td>Bachem et al., 1996; Habu et al., 1997</td>
<td>Bauer et al., 1993; Bertioli et al., 1995; Wan et al., 1996; Poirier et al., 1997</td>
<td>Fischer et al., 1995</td>
<td>Sambrook et al., 1989</td>
<td>Wan et al., 1996</td>
</tr>
</tbody>
</table>

Reproducibility of cDNA-AFLP and DDRT-PCR: number of differential bands visible in different experiments/number of bands examined (%). Sensitivity is given in terms of mRNA abundance (mass of desired mRNA/mass of total mRNA in RNA probe). The size value for subtracted libraries assumes ten-fold enrichment of the differentially expressed cDNAs by the subtraction step. Total RNA is for one experiment assuming no amplification. DS, Differential screening; SH, subtractive hybridization; PAGE, polyacrylamide gel electrophoresis; CE, capillary electrophoresis.
WHICH METHOD SHOULD BE USED?
The answer depends on several factors including the type of problem to be solved, the amount of sequence information available for the species under study and the abundance of the transcripts that one is looking for. In many cases choice of methods is likely to be dictated by cost. The investment in time and money needed to establish a quantitative high throughput display system is still beyond the scope of many laboratories. Following the branches of the binary tree (Fig. 5), one should consider that different methods may have overlapping fields of application, and hence clear cut decisions may be difficult.

COMMON DRAWBACKS OF AVAILABLE DISPLAY METHODS
All the methods discussed so far produce a very large amount of redundant and/or insignificant data. In part this results from one mRNA species generating more than one signal. The number of clones or elements in a representative cDNA library or cDNA array produced by arraying anonymous cDNA clones exceeds by more than an order of magnitude the complexity of the mRNA population from which the cDNA was synthesized (Table 1). If cDNA-AFLP is performed with one enzyme, each cDNA species should be represented by several bands (Habu et al., 1997).

Display System Selection Basics

**Intention to measure quantitative differences in gene expression**

- Costs limiting
  - 3' end sequences of target transcripts known
  - no sequence information available
    - anonymous clones available, mRNA abundance > 0.01%
    - No clones available or mRNA abundance ≤ 0.01%
  - Costs not limiting
    - Sequence information available
      - Target genome sequenced
      - Target genome not sequenced, sequenced cDNA clones available
      - No sequence information available

- **Intention to detect merely qualitative differences in gene expression or to isolate differentially expressed genes**
  - Sequence of a FSD known
    - No sequence information available
      - Sequencer or capillary electrophoresis available
        - mRNA abundance < 1%
        - mRNA abundance ≥ 1%
      - No high resolution separation technique available
        - mRNA abundance > 0.05%
        - mRNA abundance ≤ 0.05%

Fig. 5. Flow chart summarizing a few rules of thumb that may facilitate choice of the most appropriate display system for a particular application. DS, Differential screening; FSD, family specific domain; SH; subtractive hybridization.
In AFLP the use of amino-blocked AFLP adapters results in only the 3' end fragments being amplified. Hence each cDNA generates no more than one band. By combining this approach with recent cDNA amplification protocols (Matz et al., 1999) it might be possible to completely remove redundancy from the traditional cDNA-AFLP procedures. DDRT-PCR is inherently redundant, since a given mRNA is usually recognized by more than one pair of primers and hence generates different bands in different lanes of the gel during separation.

A more prominent source of meaningless data arises from the fact that usually only a few genes are expressed differentially under different conditions. For example, cDNA microarray analysis of 673 genes from two Arabidopsis accessions revealed that expression levels of only two genes differed remarkably between the two strains (Kehoe et al., 1999). SAGE analysis of approx. 1500 genes from aerobically and anaerobically treated rice seedlings showed that only 24 transcripts were expressed at significantly different levels (Matsumura et al., 1999). Elimination of the common background produced by constitutively expressed genes constitutes a major problem. In principle this can be solved by algorithms designed to eliminate all insignificant differences in signal strength, thereby retaining only those exceeding a certain threshold. However, the necessity to generate and to process expression signals from hundreds or thousands of genes not directly involved in the biological process under study remains. Processing of gene expression data would be substantially simplified if there were a method of biochemical subtraction allowing only those transcripts to be recorded that are differentially expressed. Subtractive hybridization is the only method developed so far to achieve this. The results, however, have been unsatisfactory.

All differential display techniques provide information about steady state mRNA concentrations and cannot detect post-transcriptional events known to play a role in plant gene regulation such as changes in mRNA stability, translation efficiency or protein (de)phosphorylation.


