– Review paper

Plant Tissue Cult. **12**(2) : 195-213, 2002 (December)

PTC

Gene-chips in Plant Genetics

Narayana R. Isola*, A. Arundhati[1], T.V. Rajya Lakshmi[1] and A. Aruna Lakshmi[1]

Oak Ridge Biotech (OrBit), 148 Brentwood Circle, Oliver Springs, Tennessee 37840, USA. E-mail: nrisola@aol.com

Must we geneticists become bacteriologists, physiological chemists and physicists, simultaneously with being zoologists and botanists? Let us hope so.
– H. J. Muller

Abstract

Recent developments in the microarray technologies have contributed substantially to our understanding of gene regulation in plants. Availability of complete genome sequences coupled with the microarray methods allows a global comparison of the gene expression profiles and the development of a transctiptome. Transcriptome is a catalogue of all the expressed messages present in a cell (tissue). This approach, although in its development stages, has already proved to be invaluable in identifying new genes involved in signaling pathways as well as in confirming earlier findings in the elucidation of these pathways. This article is aimed at reviewing state-of-the-art of microarray techniques. A summary of the diverse applications of microarays in plant genetics is provided with suggestions for future research.

Introduction

Genetics in general and plant genetics in particular originated a little over a century ago with the rediscovery of Mendelism by Correns, Tshermark, and DeVries. The 20th century has witnessed several major advances in this field that not only revolutionized the field of genetics but also influenced every branch of biology and medicine. The borderline between various disciplines in biology is fast disappearing and new fields such as biotechnology and bioinformatics have emerged. The milestones and breakthroughs in this area are too numerous to mention and is not the subject of the present review. "DNA makes RNA and RNA makes protein" (Watson and Crick 1953) (Fig. 1a) was the central dogma as initially proposed. However, the discovery of reverse transcriptase changed the central dogma as shown in Fig. 1b. The central dogma holds true for all living organisms with minor variations such as differences in the genetic code that was once considered universal, discovery of introns and

---

*Author for correspondence. [1]Department of Botany, Andhra University, Waltair, 530 003, India.

genes coding for multiple peptides and so on. The source of genetic information is in the genome (typically DNA) that is transcribed into RNA and translated into proteins. The genome remains constant in all the tissues and at all times in an organism, while the RNA and protein repertoire and concentrations vary considerably. Understanding gene expression is the main focus of biological research currently.

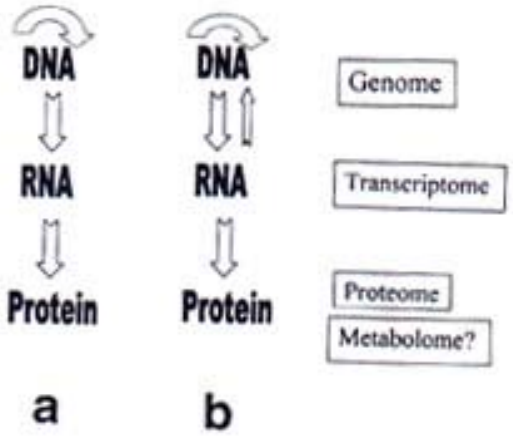

Fig. 1. The central dogma as initially proposed by Watson and Crick (a) and later modified to include reverse transcription and RNA viruses (b). The current thinking of the sum of all the genes (genome), all the transcripts (transcriptome), all the proteins (proteome) and all the secondary metabolites (metabolome) is shown alongside.

The last quarter of the 20th century has truly revolutionized genetics and biology in general. This was partly due to the enormous insurgence of physical scientists into the realm of biological research. Several techniques were developed in order to understand the details of cellular machinery. Discovery of restriction/modification enzymes (Smith and Wilcox 1970), and use of these enzymes for *in vitro* manipulation of nucleic acids (Cohen et al. 1972) initiated the molecular biology revolution. Methods for detection and analysis of nucleic acids based on sequence specific hybridization (base pairing) (Southern 1975) and the development of methods for sequence determination (Sanger et al. 1977, Maxam and Gilbert 1977) have carried the revolution further in our understanding of genes and their functions in various organisms. Methods for *in vitro* amplification of specific sequences (RNA and/or DNA) were developed less than two decades ago. Starting with the original polymerase chain reaction (PCR) (Saiki et al. 1988), subsequently followed by other methods such as self sustained sequence reaction (3SR), Ligase Chain Reaction (LCR), QB

replicase and so on (Carrino and Lee 1995). These techniques have been instrumental in developing the new molecular diagnostics field.

The above mentioned technologies have led to the crowning achievement of the 20th century deciphering the complete nucleotide sequence of the human genome. Several other viral (lambda phage, M13, HIV), bacterial (*E. coli*, *Hemophilus*), and eucaryote (*Coenorhabditis elegans*, *Saccharomyces cereveciae*) genomes have been sequenced initially as model systems. The techniques are currently used for sequencing a variety of organisms. The complete nucleotide sequence of *Arabidopsis thaliana* has been released (AGI-Arabidopsis Genome Initiative 2001) and the *indica* (Yu et al. 2002) and *japonica* (Goff et al. 2002) varieties of rice sequences have been determined. These should serve as model systems for monocots and dicots and a complete understanding of the gene regulation at the global scale could be expected in the near future.

Development of the macroarray technologies

Gillespie and Speigelman (1965) demonstrated that single stranded DNA binds strongly to nitrocellulose membranes in such a fashion that would allow complementary sequences to be used as probes to analyze the bound sequences. This method has been the cornerstone in the development of DNA reassociation kinetics analysis and the discovery of unique and repetitive sequences in higher organisms. This method has been further developed by Southern (1975) into what is popularly known as Southern blot. The above procedure was instrumental in understanding the genome structure of several genes and several modifications of the original method soon followed. One such important variation is the dot blot in which DNA/RNA from various sources was spotted onto a membrane and probed with a radiolabeled probe. This allowed for screening thousands of samples simultaneously for the presence of a specific target sequence. An important variation called the reverse dot blot was later developed. In this method the probe (known sequence) sequence was immobilized on to the substrate and target (analyte), generated *in vitro* (typically by PCR amplification) with appropriate labels and hybridized to the membranes containing the probe sequences. The other methods typically immobilized the unknown or analyte sequence on to the membrane and are probed (known sequence). In dot blot method the presence of a single probe can be analyzed in hundreds of samples simultaneously, while in the reverse dot blot format the presence of several sequences in one analyte can be analyzed. This important variation was the forerunner of the present microarray

technology. Currently two different types of microarrays are available commercially : Oligonucleotide arrays and cDNA arrays.

Oligonucleotide arrays : Short synthetic oligonucleotides of defined sequences are attached covalently to a solid surface (glass, plastic, nylon etc). This is achieved by two different methods, 1. Direct (Parallel) synthesis (*in situ* synthesis) or 2. Spotting.

Parallel synthesis (Photolithography) : Parallel synthesis of high-density oligonucleotide arrays is achieved by photolithographic methods. The most widely used method developed by Fodor et al. (1991). The surface is initially derivatized with chemical linkers containing photolabile groups. A light mask is deposited onto the surface and selective regions are activated, by shining light through the mask. The surface is covered with modified nucleotide to be coupled. After coupling the initial mask is removed and a new mask is printed on the surface and the cycles repeated. This selective activation of specific regions and coupling of specific nucleotides allows parallel synthesis of oligonucleotides. This is also known as the Affymetrix

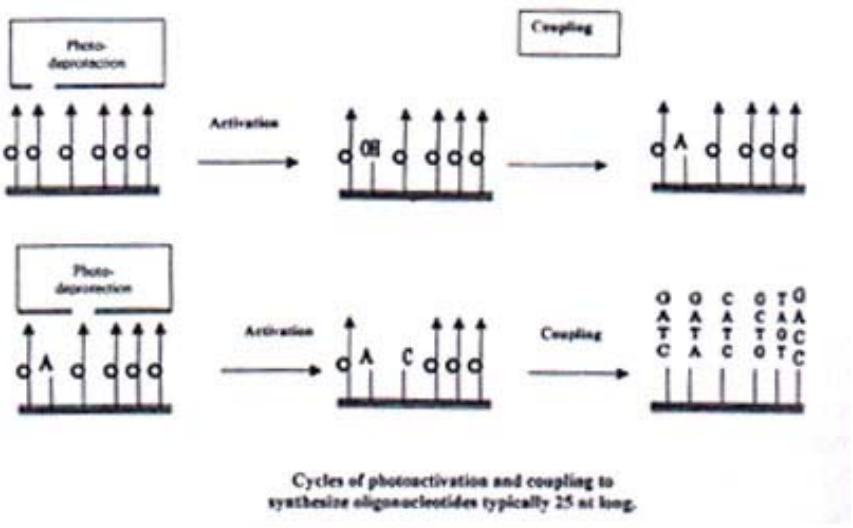

Fig. 2. Schematic representation of the photolithographic process of parallel synthesis of oligonucleotide arrays. The first step of masking the slides and photoactivation of selected regions followed by coupling of the nucleobase to the activated regions. The pattern for masking and photoactivation and synthesis are computer controlled and the cycle is repeated to generate oligonucleotide probes 25 - 30 bases long typically.

chip and it typically contains 25-mer oligonucleotides within 20-25 μm spot. This method is schematically shown in Fig. 2. Current advances in Nanotechnology are aimed at generating spots in the nanometer range so that the complete genome of an organism can fit onto a single slide.

Spotting

In this method modified oligonucleotides (typically amino linker or thiol linker) are spotted onto derivatized supports either using a contact method (A modified version of the classical fountain pen nib Schena et al. 1995) or a non-contact inkjet technology (Okamoto et al. 2000). This method is more convenient and inexpensive compared to direct synthesis method for oligonucleotides. The oligonucleotide chips are variously referred to as the Affymetrix chips (based on the initial producer), genechips or sequencing chips. These chips were originally designed to obtain sequence information from short regions of DNA, also known as re-sequencing, since original sequence information is necessary for the design of the chip. This chip can also be used for obtaining SNP (single nucleotide polymorphism) data in small regions of the genome. Very few studies in plant sciences have used this type of chip.

The second and more common chip, also known as expression arrays, was initially developed at Stanford University (Schena et al. 1995); it involves the use of PCR amplified long cDNA molecules. This chip is also referred to as Stanford chip or expression array. In this method 1 - 2 kb regions of cDNAs are spotted directly onto polylysine treated glass slides. The oligonucleotide microarrays contain short sequences covalently attached at one end (often with short to medium linkers with 6-12 carbon atoms) this allows the probe sequence for free rotation in solution and allows the complete sequence to be available for hybridization. While in the cDNA microarrays double stranded nucleic acids are used, these nucleic acids are held by electrostatic interactions on the surface and it is not known if the complete sequence is available for hybridization. This is the most commonly used chip in plant expression studies and other organisms as well. The increasing availability of sequence information from various organisms has led to the development of species-specific cDNA arrays. Currently *Arabadiopsis* and *Zea mays* gene chips are commercially available and various companies are designing expression arrays for other species.

Fluorescent labels

Several different non-radioactive labels for nucleic acid analysis have been developed in order to replace the hazardous and short-lived radiolabels. Colorimetric methods using Biotin (Langer et al. 1981) and Digoxgenin (Kessler

et al. 1990) probes, Chemilumine - scent (Bronstein et al. 1990) detection methods using antibodies labeled with alkaline phosphatase and a chemiluminescent substrate, SERS (Surface Enhanced Raman Spectroscopy) probes using Cresyl fast violet (Isola et al. 1998, 2002) are among the many non-radioactive methods currently in use for nucleic acid analysis. However, the advent of fluorescent labels truly revolutionized the analysis. Fluorescent labels that were initially introduced in the development of automated DNA sequencing methods (Smith et al. 1986), have slowly replaced the radiolabels in certain types of experiments such as *in situ* hybridizations. FISH (Fluorescent *in situ* hybridization) has totally replaced ^{3}H, and ^{14}C labeling methods for *in situ* hybridizations that were previously used.

Several fluorescent reporters are currently available for incorporation into nucleic acids both during oligonucleotide synthesis as well as *in vitro* enzymatic incorporation methods. A substantial number of them are derivatives of fluorescein and Rhodamine. However, the Cyanine based dyes Cy3 and Cy5 have been used more in the microarray methods. Since the enzymatic incorporation of these dye labeled nucleotides has been optimized and the dyes have absorption maxima close to each other, allowing the use of a monochromatic laser for illumination. The emission is sufficiently distinct that the two dyes can be readily resolved using standard emission filters.

Typical expression profiling experiment

In a typical experiment RNA isolated from two samples to be compared are used. The RNA populations are converted to cDNA, using reverse transcriptase and in most experiments subjected to amplification (Fig. 3). During the *in vitro* nucleic acid synthesis Cy3 labeled nucleotide is incorporated into one sample and Cy5 labeled nucleotide is incorporated into the other. The two samples are pooled at equimolar concentrations and hybridized onto the microarray. After hybridization and stringency washes the micro-arrays are scanned using a scanner, typically equipped with a confocal microscopy.

A number of different microarrays and scanning equipment are currently available in the market and research groups are constantly upgrading the results and comparing them in order to achieve a global standard method for labeling, hybridization and scanning as well as for statistical analysis of scans. Certain agreements have been achieved and will soon be revised and updated periodically. This is the sign of the growing technology; and collaboration between researchers working in these areas, would help achieve a standard method of comparison of results from one laboratory to the other as well as facilitate easy methods for data mining and interpretations.

Although the microarray technique is less than ten years old, the number of laboratories (consortia) using these methods is increasing exponentially and already several hundred publications in plant genetics alone have been published, utilizing the results from microarray analyses.

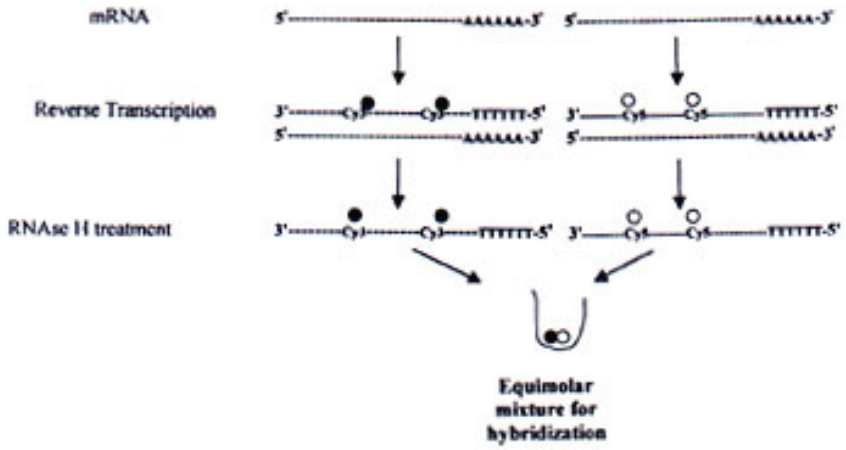

Fig. 3. Messenger RNAs from two samples are converted to cDNAs using reverse transcriptase and labeled with either Cy3 or Cy5 during the *in vitro* synthesis process. The RNA molecules in the DNA-RNA hybrids are degraded using RNAseH and the resulting single stranded mixtures are used in equimolar ratio to hybridize single stranded oligonucleotide arrays. Agilent Technologies.

Future

Microarrays on solid substrates were initially conceived as an alternative to blots for the following reasons :

(a) The availability of high efficiency fluorescent labels and laser scanners, a high density of probes can be achieved in a small space, thus saving a large number of expensive reagents.

(b) Attachment of probes to solid surface would cut down the hybridization time considerably, since there are no diffusion-related problems as in the porous membranes. Vo-Dinh et al. and Isola (unpublished) have demonstrated that on a solid surface such as plastic or glass, hybridization equilibrium can be achieved in min, when single-stranded target and probe molecules are used. This is similar to the time observed in solution hybridization experiments. However, the microarray literature clearly shows that hybridizations are typically performed overnight (about 16 h). Methods

for converting double stranded molecules into single stranded forms will have to be optimized for use both in the cDNA microarray preparation as well as in the sample preparation for both oligo arrays and expression arrays in order to achieve fast hybridizations as mentioned above. Two important variations are already implemented in the Agilent microarray.

1. Use of single stranded target sequences without the use of amplification technology.

2. Use of single stranded probe sequences derived from the cDNA sequences. It has been well established that the use of full-length cDNAs as probes occasionally leads to ambiguous results due to the repeat sequences present in cDNAs.

Plant genetics

Since the pioneering work on garden peas (*Pisum sativum*) by Gregor Mendel that led to the elucidation of basic principles of genetics, plant systems were used as experimental models to address various questions related to inheritance. Plants were chosen as model systems as well as for their economic importance. The first known interspecific gene transfer was achieved by transferring the rust resistance gene from rye to wheat, using classical genetic techniques several decades prior to the development of *in vitro* recombinant DNA technology (Sears and Okamato 1958). Triticale (hybrid of wheat and rye) is the only synthetic species to date with commercial applications. Elegant genetic experiments and deductive logic led to the discovery of transposons in corn by Barbara McClintock. Principles of cytoplasmic inheritance were determined by Dr. Ruth Sager using the chloroplasts of Chlamydomonas as model systems. These are but a few examples of the contribution of plants to the fast growth and development of general genetics. Although human- and animal genetics have become technology intensive and of interest due to the immediate applications, the power of the genetic crosses makes plants better systems for genetic analysis. Initial observations and analyses of the progeny lead to model building and the models can be readily tested using appropriate back- and test crosses. In animals and humans, this is not possible, forcing the use of alternative methods such as somatic cell genetics, reverse genetics etc.

Microarrays in plant genetics

Expression of genes is broadly classified as constitutive and inducible. Constitutive genes typically are genes whose functions are required for basic

biochemical mechanisms in the cell such as the TCA cycle, glycolysis, and pentose phosphate pathway. Inducible genes are genes that respond to external and/or internal stimuli such as light, chemicals, and stress. Although this classification is not rigid a comprehensive survey of gene expression in model systems would help redefine this classification and provide an understanding of different pathways that respond to different stimuli. The goal of the global expression studies is to identify and understand the biochemical pathways and their interconnections in model plants. New methods are constantly being developed from the information and tools developed so far. The genomic revolution is leading towards the development of functional genomics, proteomics, metabolomics and so on. With this trend we might be able to trace all the biochemical pathways and their interrelations in some model systems in the near future. Even though the expression profiling method is still in its infancy the use of this technique has already been fruitful in understanding a number of aspects of plant genetics.

For the purposes of this review the various classes of research are classified into the following groups.

1. Developmental gene regulation.
2. Tissue specific gene expression.
3. Gene expression in response to abiotic factors.
4. Gene expression in response to biotic factors.

1. Developmental gene regulation

Jian-Ming Lee et al. (2002) dissected these metabolic pathways and their interactions in maize embryos using expression profiles. Temporal patterns of gene expression during the development for genes involved in carbohydrate, fatty acid, and amino acid metabolism, tricarboxilic acid cycle, glycolysis, the pentose phosphate pathway, membrane transport, signal transduction, cofactor biosynthesis, photosynthesis, oxidative phosphorylation and electron transfer. These studies provide a valuable database and reference point in order to understand the interrelations of these pathways in response to other stimuli. These results were consistent with published data. For example, genes involved in DNA synthesis and in protein synthesis produced high steady-state levels of mRNA during early embryo development, with a decline observed in the late stages of embryo development. The genes encoding enzymes involved in the TCA cycle and glycolysis are coordinately regulated. The expression of TCA cycle

enzyme genes is highly coordinated, that is all temporal patterns of expression are very similar, while glycolysis appears to show more differences in temporal patterns of expression. This may reflect different regulatory pathways and/or involvement of these enzymes in other functions in the cell. Once the global patterns of expression is established in growth and development, isolation of genomic clones and the promoters associated with these genes would provide information regarding such coordinated expression as demonstrated in yeast (Cho et al. 1998).

Fig. 4. *Arabidopsis* microarray. Each spot consists of single stranded 60 mer. These 60 mer sequences are derived from the unique sequences in the cDNAs. 3,620 spots/cm^2. Each microarray contains 90 repetitions of the control probe. (The series of dots in the corners).

With the availability of completed *Arabidopsis* genome sequence (AGI 2001) consisting of a total of 25,498 identified genes using gene prediction software, a truly global screening of the genome is possible. Zhu et al. (2001) utilized this piece of information and studied 8,300 genes in microarray format for their expression in various tissues of the plant (Fig. 4). They have identified 347 genes as constitutive. Based on their functions, the constitutive

genes have been classified as follows, the percentage of genes involved in each category, being shown against each class in parenthesis : metabolism (33), ribosomal protein synthesis (15), transcription (8), signaling (6), transport (5), membrane (5) and stress and defense related (7). No function could be assigned to about 15% of the identified genes.

2. Tissue specific gene expression

Differentially expressed genes are classified based on their organ and development specific expression. Number of genes expressed differentially in root (64), leaf (94), inflorescence stem (3) and flower (36) were identified and functionally characterized. Identification of co-regulated genes is the first step towards understanding pathways and their mechanisms of regulation. However, assigning functions to the vast number of orphan genes is the primary task of the genome project. A major component of human diet consists of fruits. Fruits provide a number of vitamins, minerals, and antioxidants and are a good source of energy. However, the flavor of the fruit is of utmost importance. Flavor of fruits is due to a number of secondary metabolites produced during the ripening process. Biochemical study of fruit ripening processes has yielded a number of benchmarks as guidelines for geneticists and breeders for development of improved varieties of fruits. Strawberries *(Fragaria grandiflora)* contain more than 300 different compounds in the ripened fruit that contribute to the flavor. These compounds are grouped into several classes, comprising acids, aldehydes, ketones, alcohols, esters, lactones *etc*. Esters are one of the most important compounds in fruit flavor and in strawberry alone, more than a hundred different esters have been detected (Honkanen and Hirvi 1990). The enzyme catalyzing the esterification reaction is termed alcohol acetyltrans-ferase (AAT), a key enzyme in aroma biochemistry. Aharoni et al. (2000) used cDNA microarray analysis of mRNAs from developing fruits of strawberry and identified the strawberry alcohol acetyltransferase (SAAT) enzyme. Detailed RNA blot analysis experiments confirmed that the first expression of SAAT was during the white stage of fruit development and it is expressed exclusively in the receptacle tissue of the fruit.

Most angiosperms produce hermaphrodite flowers. It is of great importance to understand the developmental mechanisms, involved in the male and female organs of the flower as well as the triggers that turn vegetative growth into reproductive stage. Two major strategies have been used to identify tissue specific genes. Genes involved in sexual reproduction have been isolated

using subtractive hybridization, differential plaque hybridization, and differential display. Genes have also been identified based on the differences in protein profiles and micro-sequencing. However, not all members of the organ specific genes have been identified as yet. It is estimated that a pollen grain contains about 2000 specifically expressed genes (Willing et al. 1988). Endo et al. (2002) used microarray methods to study the model legume *(Lotus japonicus)* and identified 111 independent cDNA clones, that were expressed in the mature anther. They have similarly isolated clones specifically expressed in the immature anther and the pistil as well. Sequencing information revealed that a majority of these sequences are unique.

3. Gene expression in response to abiotic factors

Several physiological responses in plants are adapted to the diurnal cycling of light and dark. Plants respond to light through photoreceptors. In *Arabidopsis*, two classes of photoreceptors are known. Phytochromes A through E responds to red/far-red regions of the spectrum while other receptors (CRY1, CRY2, NPH1) respond to the blue light. Schaffer et al. (2001) used microarrays and discovered that out of the 11,521 clones used as probes, the expression of 816 of them was in a diurnal pattern and 139 were in a circadian manner. Many of these genes have not been previously reported to be regulated by light. This demonstrates the power of this technology. With this piece of information further analysis of these genes will allow the understanding of the primary and secondary factors, involved in the circadian cycles in plants. Further analysis will provide information regarding these genes that are regulated by the circadian clock and those that function as the clock.

Plants undergo several environmental stresses such as cold, heat, salinity and the responses of different plants to these abiotic stresses is different. An understanding of the response mechanisms may prove to be a key factor to tailor crop plants to grow in a variety of environments. Seki et al. (2001) studied drought and cold stress induced genes in *Arabidopsis thaliana,* using microarray technology and identified 44 drought- and 19 cold-inducible genes. Among them 30 and 10 are novel drought- and cold-inducible genes, respectively. This study consisted of about 1300 full length cDNA clones as probes and the researchers are currently planning to construct arrays, consisting of greater than 10,000 full length cDNAs to get a complete listing of the genes that are stress inducible. This research should provide most if not all the genes directly and indirectly involved in stress responses; and sequencing and expressing these identified genes would provide insights into the mechanism of stress response by *Arabidopsis thaliana.*

As mentioned above since plants survive and respond to different types of environment, stress conditions for one species may provide normal habitat to another. This makes it very difficult to generalize stress responses. However, the microarray technology can provide basic outlines fairly rapidly that can eventually be fitted in a global picture.

Kawasaki et al. (2001) studied genes responding to salt stress in rice. They compared a salt tolerant variety Pokkali with the salt intolerant variety IR29. The expression profiles in Pokkali were further categorized into several patterns using cluster analysis (Eisen et al. 1998). The clusters are represented by transcripts that are upregulated at different times; 15 min, 1, 3 and 6, 24 hr and 7 days. A distinguishing feature, revealed by this microarray analysis, was the timing of otherwise similar responses. The expression profiling Pokkali after 15 min treatment was similar to that seen for IR29 after 1 h, suggesting that a delay in the processing of signals could account for ineffective response of IR29 to salt stress. Analysis of more strains of rice using this type of experiments would help pinpoint the underlying gene expression changes that are the basis of a plethora of physiological responses and help in breeding salt tolerant varieties rapidly.

4. Gene expression in response to biotic factors

Use of global gene expression analysis method of the plant gene expression as well as the pathogen gene expression is already beginning to yield valuable information. This body of information will be invaluable in designing control strategies either through genetic, cultural practices, or through chemical means. Thanks to the green revolution during the 1950's, a wide number of high yielding varieties of crop plants such as wheat, rice were developed and the global production of grains and food crops has nearly doubled. It is estimated that nearly 25% of the food never reaches the market due to insect attacks and virulent diseases?

Plant pathogen interactions

Understanding the plant pathogen interactions is very important in order to be able to control the wide variety of plant pests on crop plants. Thanks to the advances during the green revolution, the productivity of crop plants has increased tremendously. However, the full benefit of these high yielding plants cannot be realized without appropriate control of the diseases and pests that these plants are susceptible to.

A number of research articles describe the use of microarrays for the understanding of plant pathogen interactions. Reymond (2001) reviewed some of the work from *Arabidopsis*. In this review we would like to highlight some of the salient features of different discoveries.

Cheong et al. (2002) surveyed 8200 genes in *Arabidopsis* using microarrays for their response to wounding, pathogen, abiotic stress and hormonal responses. They discovered a substantial number of genes that respond to wounding, which also responded to pathogen interaction. Wounding and pathogen responses led to stimulation of the number of hormones such as jasmonic acid, and ethylene. These hormones in turn induced the expression of a number of defense genes such as proteinase inhibitors, thionin and enzymes involved in secondary metabolism. Perhaps the most surprising finding in this study was the interaction between wounding and auxin signaling pathway. This study revealed a new cross talk between the various biochemical pathways in the cellular response to external stress.

Defense response in plants depends on the ability of the host to recognize the pathogen and initiate the defense reactions. These reactions can be classified as systemic (or general) and specific (specific to the pathogen). Production of a wide variety of antimicrobial peptides and structural modifications have been observed as well. Elucidation of the endogenous signal molecules such as salicylic acid, jasmonic acid and ethylene have been demonstrated and are being studied in various plants. Schenk et al. (2000) analyzed coordinated defense response in *Arabidopsis* using microarrays and discovered that 55 genes were co-induced by both salicylic acid and jasmonic acid. This contradicts the previous notion that these two pathways are antagonistic. However, the results also show that eight genes were significantly induced by salicylic acid and significantly repressed by methyl jasmonate indicating that signal antagonism may be specific for specific genes.

Most of the above mentioned studies were performed using only partial genomes and are already contributing significantly to our understanding of gene regulation. A significant understanding of the different pathways and especially the cross talk between them would be extremely beneficial in the design of both resistant varieties as well as other pathogen control methods such as altered cultural practices, pesticides.

Future

From the biological perspectives plant systems have a considerable variety of systems to offer that would benefit from microarray analysis.

1. B chromosomes : The standard set of chromosomes is referred to as 'A' chromosomes and supernumerary or extra chromosomes are referred to as 'B' chromosomes. A number of plant species have been shown to have varieties consisting of these B chromosomes in variable number(s)/cell. An increase in the plant growth and vigor has been observed in plants containing a low number of B chromosomes (Moss 1966, Stebbins 1971). Since these chromosomes do not contain a defined centromere the segregation pattern of B chromosomes is not as predictable as the normal chromosomes and it is possible to breed plants containing the desired number of B chromosomes. This type of system would be invaluable for microarray analysis in order to understand both the genetic compositions of the B chromosomes as well as the influence of the B chromosomes in transcription. The segregation behavior of these B chromosomes is quite well characterized in plants such as maize, rye, *Pennisetum* and A-B translocations have been isolated as well. These translocations would be of great importance in transferring desirable traits from the B chromosomes into cultivated varieties. Thus, these B chromosomes are not only of basic biological interest but it would also provide a novel source, allowing manipulation of genetic material inside the cell to accelerate plant breeding programs . The pairing suppressor gene (Ph) that suppresses (Riley and Chapman 1958) the pairing between homoelogous chromosomes without affecting the pairing of homologues has been shown to be derived from B chromosomes. A clear molecular mechanism of this fascinating phenomenon would certainly advance our understanding of meiosis.

2. Polyploidy : The polyploid series of plants are a valuable system to study gene regulation and interdependency of the different genomes. An understanding of gene regulation in this series would serve as model system for other polyploid series such as *Brassica, Solanum* and so on. Increased understanding might even provide clues to produce desired combinations using *in vitro* techniques in the future as well.

3. Totipotency : Stewart (1963) discovered the totipotent nature of the plant cell using carrot disks. Certain plant cells that are differentiated into mature organs can under certain conditions revert to the meristematic nature and re-initiate differentiation. This is the basis of plant tissue culture. A clearer and comprehensive understanding of this phenomenon can result using complete cDNA arrays and this would certainly be beneficial for the *in vitro* propagation of crop plants and clonal preservation.

Somaclonal variation or tissue culture variation has become an established fact in almost all the plant populations subjected to *in vitro* culture.

The mechanisms by which this variation occurs and how to control this variation are still not clear. A global genome comparison and transcriptome comparison in model systems would certainly shed light on the mechanism of somaclonal variation. It is also quite possible that the genome is not the same in all tissues as it was thought to be. If one imagines 109 base pairs of the DNA replicating to produce an organism with $10^{14\text{-}16}$ cells, it would mean differentiated tissues would accumulate a substantial number of base changes, assuming an average mutation rate of 10^{-7}/ base pair. Coupled with the activity of transposons in several plant species the complete genome sequence is likely to be different in different tissues.

4. PTGS : Post transcriptional gene silencing was initially encountered in plants during attempts to over-express endogenous genes by the introduction of sense constructs (Ann Depicker and Marc Van Montagu 1997). It is now documented in insects, nematodes and mammalian cells. With the wide variety of genetically engineered plants available, these will be suitable for microarray analysis to get a comprehensive understanding of this intriguing biological phenomenon.

5. Meiosis : The most interesting phenomenon in eucaryotes is the transition of cells from mitosis to meiosis. A complete understanding of the triggers that switch cells from mitotic- to meiotic cycle would be of great biological interest. With potential applications in infertility, cloning etc., Yeast has been the model system to study cell cycle and a large body of information on different genes that act coordinately have been identified (Honigberg et al. 1991). Plants especially the male gametophyte (pollen development) offer a highly synchronous population of cells in various premeiotic and meiotic stages. As the genomes of rice and *Arabidopsis* are available, it would be interesting to analyze these cells at various stages and in various meoitic mutants to gain a comprehensive understanding of this fascinating biological transition.

Acknowledgements

The authors wish to thank Agilent Technologies, Palo Alto, CA, USA for sharing the Arabidopsis microarray images. They also wish to thank all the researchers who gladly responded in sharing the published (and unpublished) work through E-mails. They wish to thank Mr. Y. Satyanarayana for computer help. The first author (NRI) wishes to thank Mr. David Hughes for critical review of the manuscript. In an effort to summarize work that is fairly new as

well as broad ranged errors of omission are expected and they apologize for not being able to cite some other excellent work in this area.

We dedicate this article to the memories of Prof. J. V. Pantulu and Prof. M. V. Krishna Rao who have been our inspirational teachers that introduced us first to Mendelian and modern genetics.

References

Arabidopsis Genome Initiative (AGI) 2001 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* Nature. **408** : 796 - 815.

Aharoni A, Keizer LCP, Bouwmeester HJ, Sun Z, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen AMML, De Vos RCH, van der Voet H, Jansen RC, Guis M, Mol J, Dawis RW, Schena M, van Tunen AJ and **O'Connell AP**. 2000 Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. The Plant Cell **12** : 647 - 661.

Ann Depicker and **Marc Van Montagu** (1997) Post transcriptional gene silencing in plants. Curr. Opin. Cell Biol. **9** : 373 - 382.

Bronstein I, JC Voyta C, KG Lazzari G, Murphy O, Edwards B, Kricka LJ (1990) Rapid and sensitive detection of DNA in Southern blots with chemiluminescence. Biotechniques **83** : 310 - 4.

Carrino, JJ and **Lee HH** (1995) Nucleic acid amplification methods. J. Microbiol. Methods, **23** : 3 - 20.

Cheong YH, Chang H-S, Gupta R, Wang X, Zhu T and Luan S (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. Plant Phys. **129** : 661 - 677.

Cho RJ, Campbell MJ, Winzler EA, Steinmetz L, Conway A, Wodicka L and **Wolfsberg TG.** (1998) A genome-wide transcrptional analysis of the mitotic cell cycle. Mol. Cell. **2** : 65 - 73.

Cohen SN, Chang ACY, Chang HW and **Boyer HW** (1972) **Co**nstruction of biologically functional bacterial plasmids *in vitro.* Proc. Natl. Acad. Sci. USA. **70** : 3240.

Eisen MB, Spellman PT, Brown PO and **Botstein D** (1998) Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci., USA. **95** : 14863 - 14868.

Endo M, Matsubara H, Kokubun T, Masuko H, Takahata Y, Tsuchiya T, Fukuda H, Demura T and **Watanabe M** (2002) The advantages of cDNA microarray as an effective tool for identification of reproductive organ-specific genes in a model legume *Lotus japonicus.* FEBS Letters **514** : 229 - 237.

Fodor SPA, Read JL, Pirrung MC, Stryer L, Lu AT and **Solas D** (1991) Light-directed, spatially addressable parallel chemical synthesis. Science **251** : 767 - 773.

Goff, S. A. et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp japonica). . Science **296** : 92 - 100.

Gillespie D and **S Spiegelman** (1965) A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane, J. Mol. Biol. **12** : 829.

Honigberg SM, McCarrol RM and **RE Esposito RE** (1991) Rgulator mechanisms in meiosis. Curr. Op. Biol. **5** : 219 - 225.

Honkanen E and **Hirvi T** (1990) The flavour of berries. In Food Flavours. ID Morton and A J Macleod (eds), Amsterdam : Elsevier Scientific Publications, pp. 125-193.

Isola NR, Stokes DL, Vo-Dinh T (1998) A SERS genosensor for HIV detection. Anal Chem **70** : 1352 - 1356.

Isola NR, Stokes DL and **Vo-Dinh T** (2001) Novel non-radioactive probes for simultaneous detection of HIV and HCV sequences using SERS spectroscopy. Nucl Acids Res. (in press).

Isola NR, Stokes DL and **Vo-Dinh T** (2002) Novel non-radioactive probes for simultaneous detection of HIV and HCV sequences using SERS spectroscopy. Nucl. Acids Res. (in press).

Jian-Ming Lee, Williams ME, Tingey SV and **Rafalski JA (**2002) DNA array profiling of gene expressino changes during maize embryo development. Funct. Integr. Genomics **2** : 13 - 27.

Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kiyoshi K, Galbraith D and **Bohnert HJ** (2001) Gene expression profiles during the initial phase of salt stress in rice. The Plant Cell **13** : 889 - 905.

Kessler C. et al. (1990) Non-radioactive labeling and detection of nucleic acids.I. A novel DNA labeling and detection system based on digoxigenin: anti-digoxigenin ELISA principle (digoxigenin system). Biol. Chem. Hoppe-Seyler **371** : 917 - 927.

Langer PR, Waldrop AA and **Ward DC** (1981) Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. Proc Natl Acad Sci U S A **78**: 11 6633-7.

Maxam AM and **Gilbert W** (1977) A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74** : 560.

Moss JP (1966) The adaptive significance of B chromosomes in rye. Chromosomes Today **1** : 15 - 23.

Okamoto T, Suzuki T and **Yamamoto N** (2000) Microarray fabrication with covalent attachment of DNA using bubble jet technology. Nature Biotechnol. **18** : 438 - 441.

Reymond P, Weber H, Damond M and **Farmer EE** (2000) Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. Plant Cell **12** : 707 - 719.

Riley R and **Chapmann V** (1958) Genetic control of cytologically diploid behaviour of hexaploid wheat. Nature **182** : 713 - 715.

Schaffer R, Landgraf J, Accerbi M, Simon VV, Larson M and **Wisman E** (2001) Microarray analysis of diurinal and circadian-regulated genes in Arabisopsis. Plant Cell. **13** : 113 - 123.

Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Sommerville SC and **Manners JM** (2000) Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proc. Natl. Acad. Sci. USA **97** : 11655-11660.

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Caminci P, Hayashizaki Y and **Shinozaki K** (2001) Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. The Plant Cell **13** : 61 - 72.

Saiki RK, DH Gelfand S Stoffel SJ Scharf R Higuchi GT Horn KB Mullis and **HA Erlich** (1988) Primer directed enzymatic amplification of DNA with thermostable DNA polymerase. Science **239** : 487.

Sanger F, S Nicklen and **AR Coulson** (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. **74** : 5463.

Schena M, Shalon D, Davis RW and **Brown PA** (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270** : 467-470.

Sears ER and **Okamato M** (1958) Intergenomic chromosome relationships in hexaploid wheat. Proc. X Int. Cong. Genet. **3** : 258 - 259.

Smith LM, JZ Sanders, RJ Kaiser, P Hughes, C Dodd, CR Connell, C Heiner, SBH Kent and **LE Hood** (1986) Fluorescence detection in automated DNA sequence analysis. Nature **321** : 641 - 679.

Southern E M (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98** : 503.

Stebbins GL (1971) Chromosomal evolution in higher plants. Edward Arnold, London.

Steward FHC (1963) The control of growth in plant cells. Sci. Am. **209** : 104 - 114.

Watson JD and **Crick FHC** (1953) A Structure for Deoxyribonucleic Acid. Nature **171**: 737.

Willing RP, Bashe D and **Mascarenhas JP** (1988) An analysis of quantity and diversity of messenger RNAs from pollen and shoots of *Zea mays*. TAG **75** : 751 - 753.

Yu J et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp indica). Science, **296** : 79 - 92.

Zhu T, Budworth P, Han B, Brown D, Chang H-S, Zou G and **Wang X** (2001) Towards elucidating the global gene expression patterns of developing Arabidopsis: Parallel analysis of 8,300 genes by a high-density oligonucleotide probe array. Plant Phys. **39** : 221 - 242.