

Tissue-specific gene expression in soybean (*Glycine max*) detected by cDNA microarray analysis

Tina L. Maguire¹, Sean Grimmond², Alistair Forrest², Inaki Iturbe-Ormaetxe¹, Khalid Meksem³, Peter Gresshoff¹ *

¹ Department of Botany, The University of Queensland, Brisbane, QLD 4072, Australia

² Institute of Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia

³ Department of Plant Soil and General Agriculture, Southern Illinois University at Carbondale, Carbondale, IL 62901-4415, USA

Received June 6, 2002 · Accepted August 7, 2002

Summary

We have constructed cDNA microarrays for soybean (*Glycine max* L. Merrill), containing approximately 4,100 Unigene ESTs derived from axenic roots, to evaluate their application and utility for functional genomics of organ differentiation in legumes. We assessed microarray technology by conducting studies to evaluate the accuracy of microarray data and have found them to be both reliable and reproducible in repeat hybridisations. Several ESTs showed high levels (>50 fold) of differential expression in either root or shoot tissue of soybean. A small number of physiologically interesting, and differentially expressed sequences found by microarray analysis were verified by both quantitative real-time RT-PCR and Northern blot analysis. There was a linear correlation ($r^2 = 0.99$, over 5 orders of magnitude) between microarray and quantitative real-time RT-PCR data. Microarray analysis of soybean has enormous potential not only for the discovery of new genes involved in tissue differentiation and function, but also to study the expression of previously characterised genes, gene networks and gene interactions in wild-type, mutant or transgenic plants.

Key words: gene expression – microarray – soybean

Abbreviations: EST = Expressed Sequence Tag. – FSI = fluorescent signal intensity. – RT-PCR = Reverse transcriptase polymerase chain reaction. – TIGR = The Institute of Genetic Research

Introduction

The first plant genome was recently fully sequenced, which represents a major breakthrough in plant biology research (*Arabidopsis* Genome Initiative 2001). The information provided by the recently sequenced *Arabidopsis* genome, together with the large collection of expressed sequence tags

(ESTs) available has established a base for new approaches to studying gene expression patterns in plants. Microarray technology allows rapid gene expression analysis on a whole genome scale. Thousands of DNA fragments can be spotted at high density on a solid substrate (such as a glass microscope slide) and analysed simultaneously in a single experiment, using dual colour fluorescent probes. Gene expression profiles can then be determined over a range of experimental conditions and organised into patterns that reflect the state of

* E-mail corresponding author: p.gresshoff@botany.uq.edu.au

the tissue, or the plant as a whole. The field of whole genome expression analysis in plants is still developing, with the first report of plant microarrays by Schena et al. (1995) using 48 *Arabidopsis* cDNAs. No further plant microarray work was described until Ruan et al. (1998) reported a microarray containing 1,400 *Arabidopsis* cDNAs. Since then there have been increasing reports of plant microarrays (Schenk et al. 2000, Wang et al. 2000, Reymond et al. 2000, Yu et al. 2001, Zhu et al. 2001, Schaffer et al. 2001, Seki et al. 2001, Girke et al. 2000, Maleck et al. 2000), however most studies utilise the model plant *Arabidopsis*. There are few reports of other plant species (Aharoni et al. 2000, Kawasaki et al. 2001, Endo et al. 2002). In addition, the number of cDNAs used is still small in comparison to current applications of microarray technology in yeast and human research. This foreshadows the potential large-scale and diverse applications this technology could have in plant biology and agriculture.

The possibility of producing and using plant microarrays to study gene expression depends on sequence information from either genomic sequence data or available ESTs. World-wide numerous plant genomic research programs have been funded to carry out large-scale genomic sequencing and to establish large EST databases, and to develop microarray technology. In the near future microarray studies in many economically important crops, such as soybean, will contribute to our knowledge of fundamental questions, as well as commercially interesting traits. For example, microarray analysis could potentially identify genes involved in responses to environmental change, pathogen attack, symbiosis and many artificial treatments such as hormones and pesticides. Other applications may include linking gene expression patterns to complex traits. Using microarrays it will be possible to monitor how transgenes affect global gene expression patterns in their plant hosts, as well as gaining insights into how a gene mutation leads to a particular presence or absence of phenotype.

The large number of soybean ESTs available (>200,000) from various stages of soybean development and various organs, resulting from the Public Soybean EST Sequencing Project (Shoemaker et al. 1999), provides a basis for large-scale functional analysis of thousands of genes. This provides a rich resource for the design and production of soybean microarrays. Here we report the construction of cDNA microarrays for soybean (*Glycine max*) containing 4,089 Unigene ESTs to evaluate their application in our research on the functional genomics of nodulation in legumes. Attention needs to be paid to developing good systems for microarray analysis and verification, which will facilitate reproducible data and further application of the technology. We evaluated microarray technology by monitoring gene expression profiles in different tissue types of soybean. In addition, we evaluated the accuracy and sensitivity of the data using alternative technologies, such as quantitative real-time RT-PCR and Northern blot analysis.

Materials and Methods

Soybean cDNA clones

Soybean (*Glycine max* cv. Williams) cDNA plasmid clones of approximately 4,100 Unigene ESTs (library Gm-c1004) and various other miscellaneous clones were collected. The EST clones were generated as part of the Public Soybean EST Project (Shoemaker et al. 1999). These clones are available through ResGen, (Invitrogen Corp., e-mail: ccu@resgen.com.).

Partial sequences of the clones were analysed for sequence similarity against public protein and nucleic acid databases, and assembled into tentative consensus (TC) sequences (TIGR Soybean Gene Index, <http://www.tigr.org/tdb/gmgi/>). The total number of clones contained in the EST library is 7,480, of these 6,557 are in TCs and 923 are singletons. Of the 4,089 Unigene (clustered unique sequences) EST clones used in this study, 1,603 have no significant similarities to any coding sequences in public databases as yet, while the remainder have annotations (2,486). Visit the TIGR Current Release – Version 7.0, Release Date – November 13, 2001.

Amplification of cDNA inserts

The vector used for cDNA library construction was pBluescript II SK(+). cDNA clones were PCR-amplified from the vector using T3 (5'-AATTAACCCTCACTAAAGGG-3') and T7 (5'-GTAATACGACTCATATAGGGC-3') vector-specific primers that were modified to contain an amino group at the 5' end. Clone inserts were amplified in 100 μ L reaction volumes in 96 well PCR plates (Thermo-Fast[®]96, Abgene[®] House) using a MJ Research thermocycler as described by Hedge et al. (2000). Five microliters of each PCR reaction was electrophoresed on high-throughput (0.8%) agarose gels (Abgene[®] House) to confirm amplification quality and quantity. PCR products were isopropanol precipitated and stored dry for subsequent microarray fabrication. In general, the total quantity of each PCR product was around 2 μ g. The average size of cDNA inserts was around 800 bp.

cDNA microarray preparation

PCR products with amino groups at the 5' ends were resuspended in 30 μ L of 4 \times SSC, 0.1% sarkosyl buffer and transferred to 96-well arraying plates (Greiner 96 well Thermoquick PCR plate). The PCR products were arrayed onto amino-silane coated glass microscope slides (CMT-GAPS[™] Coated Slides, Corning Microarray Technology, Corning) using an Affymetrix[®] 417 arrayer with 250 micron spacing. A total of 4,089 Unigene EST sequences and a variety of control elements [including human and mouse Cot-DNA, several human and mouse genes including GAPDH, as well as poly(A) oligonucleotide] were arrayed in duplicate on microscope slides. Post-printing slide procedures were performed as described in Hedge et al. (2000).

Plant material and RNA isolation

Soybean (*G. max* cv. Bragg) plants were used as starting material for RNA extraction for probe preparation, Northern blot analysis and quantitative real-time RT-PCR. Plants (10 days old, first leaf stage) were grown in autoclaved sand/vermiculite (2:1) mix in a growth cabinet at 28/20 °C day/night temperature, with a 16 hour photoperiod

(500 $\mu\text{E}/\text{m}^2/\text{sec}$). Plants were watered every second day with half strength Herridge's solution, containing $5 \text{ mmol} \cdot \text{L}^{-1} \text{ KNO}_3$ (Herridge 1977). Shoot tissue (including stem, leaves and shoot tip) and root tissue (total root tissue) were collected from 5 similar plants and quickly frozen in liquid nitrogen. The pooled tissue sample was then ground to a powder by mortar and pestle, added to RLT buffer (Qiagen RNeasy[®] Midi Kit) and stored at -80°C for total RNA extraction using the manufacturers protocols (Qiagen Inc.).

Northern blot analysis

Total RNA (15 μg) was electrophoresed in (1%) agarose denaturing (formaldehyde) gels and transferred to nylon membranes (Hybond N⁺, Amersham Corp.) by capillary transfer using $20 \times \text{SSC}$ as the transfer buffer (Sambrook et al. 1989). Prior to hybridisation the nylon membranes were stained with methylene blue to confirm equal loading (Sambrook et al. 1989). cDNA probes were amplified by PCR with corresponding vector primers (M13 reverse and T7 primer). PCR products were gel purified using a QIAquick[®] gel extraction kit (Qiagen Inc.). Fifty nanograms of purified cDNA was labelled with ^{32}P dCTP using the Megaprime labelling kit (Amersham Corp.). Unincorporated nucleotides were removed using MicroSpin[™] columns (Amersham Corp.). Filters were hybridised ($0.5 \text{ mol} \cdot \text{L}^{-1}$ sodium phosphate buffer pH 7.0, 7% SDS, $1 \text{ mmol} \cdot \text{L}^{-1}$ EDTA, 0.1 mg/mL salmon sperm DNA) at 65°C overnight using standard protocols (Sambrook et al. 1989). Following hybridisation, the filters were washed under high stringency at 65°C for 15 min each using $2 \times \text{SSC}$, $1 \times \text{SSC}$, $0.5 \times \text{SSC}$, $0.1 \times \text{SSC}$ respectively, each containing 0.1% SDS. Autoradiography was performed using standard procedures.

Quantitative real-time RT-PCR

cDNA synthesis for real-time RT-PCR was as follows: total RNA (5 μg) was treated with 2U DNAase I (Gibco BRL Life Technologies Inc.) in $1 \times$ buffer in a total volume of 10 μL for 15 min at room temperature to remove DNA. One microlitre of $25 \text{ mmol} \cdot \text{L}^{-1}$ EDTA was added and incubated for 15 min at 65°C to stop the reaction. Oligo (dT) mix [5' GACCACGCGTATCGATGTCGAC(T₁₆)V 3'] was used to prime the reactions. One microlitre of $0.5 \mu\text{mol} \cdot \text{L}^{-1}$ oligo(dT) mix was added to the reaction and incubated for 10 min at 70°C , then cooled on ice. First strand mix containing $1 \times$ buffer, $10 \text{ mmol} \cdot \text{L}^{-1}$ DTT, $1.25 \text{ mmol} \cdot \text{L}^{-1}$ each dATP, dCTP, dGTP, dTTP, 40 U RNAsin (Promega Inc.) was added to a total volume of 20 μL and treated for 5 min at 42°C . Then 200 U SuperScript[™]II reverse transcriptase (Gibco BRL Life Technologies Inc.) was added and incubated for a further 55 min at 42°C . The reaction was stopped by incubating at 70°C for 15 min, then chilled on ice. To remove RNA, 2U RNase H (Gibco BRL Life Technologies Inc.) was added and incubated for 20 min at 37°C . The final reaction mix was then diluted in water to a final volume of 250 μL . Five microlitres were used for each PCR reaction. Controls (minus reverse transcriptase enzyme) were also included for each tissue type. Primers for quantitative real-time RT-PCR were designed using Primer Express[®] 1.0 software (PE Applied Biosystems) according to the guidelines for primers and probes for quantitative PCR systems (7700, 5700). PCR reactions were carried out in a total volume of 25 μL , containing $0.2 \mu\text{mol} \cdot \text{L}^{-1}$ primer (forward and reverse), $1 \times$ SYBR green PCR master mix (PE Applied Biosystems Inc.). Reactions were amplified in an ABI PRISM[®] 7700 thermocycler as follows: 95°C for 10 min, followed

by 45 cycles of 95°C for 15 sec, 59°C for 1 min, with a final incubation of 25°C for 2 min.

Preparation of fluorescent probes

Total RNA samples were reverse-transcribed in the presence of Cy3 dUTP or Cy5 dUTP (NEN, Perkin Elmer). A labelling master mix was prepared as follows: $500 \mu\text{mol} \cdot \text{L}^{-1}$ each of dCTP, dATP, dGTP, $100 \mu\text{mol} \cdot \text{L}^{-1}$ dTTP, $50 \mu\text{mol} \cdot \text{L}^{-1}$ of Cy3 dUTP or Cy5 dUTP, 40 units RNAsin (Promega Inc.), $10 \text{ mmol} \cdot \text{L}^{-1}$ DTT, $1 \times$ reverse transcription buffer. To 40 μg total RNA, 4 μg oligo(dT) 15 mer (Roche), and RNase free water were added to a total volume of 20 μL and incubated at 70°C for 10 min, then cooled on ice. To the RNA, reverse transcription labelling mix and 400 U SuperScript II reverse transcriptase (Gibco BRL Life Technologies Inc.) was added to a total volume of 40 μL . The reaction mix was incubated at 42°C for 90 min, chilled on ice, then treated with $1 \mu\text{L}$ $0.5 \text{ mol} \cdot \text{L}^{-1}$ EDTA, $2 \mu\text{L}$ $2 \text{ mol} \cdot \text{L}^{-1}$ NaOH for 10 min at 65°C to stop the reaction and degrade the RNA. Then $4 \mu\text{L}$ $1 \text{ mol} \cdot \text{L}^{-1}$ HCL and $4 \mu\text{L}$ $1 \text{ mol} \cdot \text{L}^{-1}$ Tris pH 8.0 were added to neutralise the reaction. Unincorporated fluorescent nucleotides from both reactions were removed by spin columns (QIAquick[™] PCR Purification Kit, Qiagen) according to the manufacturer's instructions. The purified products were eluted using 40 μL RNase free water, 20 μg of poly(A)-DNA (Amersham Corp.) was added and dried to a volume of 10 μL in a SpeedVac[®] (Savant Instruments Inc., USA).

Microarray hybridisation and scanning

The probe mixture was denatured at 95°C for 5 min and cooled on ice, then SSC, formamide and SDS were added to give a final hybridisation volume of 40 μL ($4 \times \text{SSC}$, 50% formamide, 0.2% SDS). The labelled probe was applied to a microarray slide and covered with a $22 \times 60 \text{ mm}$ cover slip. The slide was placed in a sealed hybridisation chamber (TeleChem International) and 10 μL of water was added to the chamber at each end of the slide. The sealed chamber was placed at 45°C in a water bath and incubated for 14–24 h. The array was then removed from the hybridisation chamber and washed with $0.2 \times \text{SSC}$, 0.05% SDS at ambient temperature until the cover slip falls off, then for a further 2 min. The slides were then washed in $0.2 \times \text{SSC}$ for a further 2 mins. The slides were dried by centrifugation. Slides were scanned with an Affymetrix[®] 418 scanner. Careful handling and pre-scanning of slide surfaces before hybridisation, as well as thorough washing steps ensured a minimum of dust and salt precipitates.

Data analysis

Stringent control measures were applied at all stages of data analysis. Control measures include: quality control of the cDNA printed on the slide by gel analysis, printing duplicate spots on each slide, performing two separate hybridisations, dye incorporation tests, and applying a ratio cut-off of at least 2 fold for differential expression. ImaGene version 4.1 (BioDiscovery Inc.) software was used for image analysis. Grids were pre-defined and manually adjusted to ensure optimal spot recognition. The quality of each spot on the microarray was assessed as follows: (1) Empty spots were automatically flagged that were not distinguishable from the background. This measures the spot and background means and calculates the difference between the two. If the difference is smaller than the threshold value (4.0), multiplied by

the background standard deviation, the spot is marked as empty. (2) Poor spots were identified based on the confidence number. If the standard deviation of the signal divided by the mean of the signal (coefficient of variance) was greater than the threshold (0.5), the spots were flagged as poor. (3) Negative spots were automatically flagged that had a signal mean lower than the background mean. Signal intensities for each spot were corrected using local background values. Data were then normalised for both channels. The intensities of the spots labelled with Cy5 were multiplied by the average of the ratio of intensity of each spot labelled with Cy3 to its corresponding spot intensity labelled with Cy5. All spots were used for normalisation of the data set. The normalised data were then log-transformed. Ratios (shoot/root) were calculated from the processed data for each channel. GeneSpring version 3.2.1 (Silicon Genetics Inc.) software was used to present the average data from both hybridisation experiments. Data were median-centred to normalise within and between slides. Gene lists were produced for all data that passed threshold values set in ImaGene 4.1 software. Because of the stringent criteria for data acceptance, it is likely that some genes that were differentially expressed may have been eliminated.

Results and Discussion

Hybridisation results

cDNA microarrays containing 4,089 soybean Unigene ESTs and internal control genes were hybridised with Cy3 and Cy5

fluorescently labelled root and shoot probes. Two separate hybridisations were performed, with two separately prepared probes from the same RNA sample applied to two copies of the same cDNA microarray. For each DNA spot the intensities (fluorescent signal intensity, FSI) of the Cy3 and Cy5 labelled probes reflect the relative abundance of the corresponding transcript in each RNA sample. Overall, about 70 % of the spots on the array generated signals above local background values. Those spots that were empty (no DNA spotted) had no visible hybridisation as expected, while spots with signal above background were visible. The hybridisation signal for spots corresponding to non-specific hybridisation controls [several human GAPDH genes, human Cot-DNA, mouse Cot-DNA, poly(A) 40mer] and negative controls (blank spots) showed background levels and did not pass threshold levels set for inclusion in the analysis.

Gene expression in shoot and root tissue

In the final data set 25 sequences were highly expressed (>10,000 FSI) in the root (Table 1) and 31 in the shoot (Table 2). There were 9 sequences that were co-expressed at a high level in both the root and shoot tissue (ratio 1.0) (Table 3). The number of sequences expressed at a modest level (between 2,000 and 10,000 FSI) was 138 and 191 in the shoot and root

Table 1. Highly abundant transcripts in root tissue.

Gene ID	Description	Ave Ratio	SD
AI441027	REPETITIVE PROLINE-RICH CELL WALL PROTEIN 1 PRECURSOR	0.003	0.001
AI441132	Novel	0.004	0.004
AI496573	REPETITIVE PROLINE-RICH CELL WALL PROTEIN 1 PRECURSOR	0.030	0.012
AI442523	FISSION YEAST	0.031	0.006
AI460620	HCAT-2A	0.044	0.024
AI440687	METALLOTHIONEIN-LIKE PROTEIN B	0.044	0.015
AI441969	SAL13-2	0.074	0.019
AI441137	ASPARAGINE SYNTHETASE 2	0.078	0.028
AI496205	ASPARAGINE SYNTHETASE 2	0.114	0.018
AI440686	MIPC	0.187	0.030
AI442288	ENDOCHITINASE PRECURSOR, BASIC	0.233	0.083
AI460606	Novel	0.270	0.032
AI496329	HYPOTHETICAL 17.7 KD PROTEIN	0.436	0.104
AI461236	PUTATIVE WATER CHANNEL PROTEIN	0.451	0.078
AI495409	Novel	0.539	0.106
AI441931	COP9	0.741	0.179
AI494859	FRUCTOSE-BISPHOSPHATE ALDOLASE, CYTOPLASMIC ISOZYME 2	0.742	0.108
AI461256	POLYUBIQUITIN PROTEIN	0.759	0.045
AI522836	ELONGATION FACTOR 1	0.962	0.141
AI440562	ASCORBATE PEROXIDASE	1.073	0.179
AI495462	ACTIN	1.162	0.209
AI460591	70 KD HEATSHOCK PROTEIN	1.312	0.130
AI495255	HISTONE H3.2, MINOR	1.328	0.156
AI496483	ELONGATION FACTOR 1	1.430	0.189
AI460407	SRC1	1.897	0.329

Table 2. Highly abundant transcripts in shoot tissue.

Gene ID	Description	Ave Ratio	SD
AI441931	COP9	0.741	0.179
AI461256	POLYUBIQUITIN PROTEIN	0.759	0.045
AI522836	ELONGATION FACTOR 1	0.962	0.141
AI440562	ASCORBATE PEROXIDASE	1.073	0.179
AI495462	ACTIN	1.162	0.209
AI460591	70 KD HEATSHOCK PROTEIN	1.312	0.130
AI495255	HISTONE H3.2, MINOR.	1.328	0.156
AI496483	ELONGATION FACTOR1	1.430	0.189
AI460551	CALMODULIN	1.446	0.101
AI440554	ADENOSYLMETHIONINE SYNTHETASE 1	1.669	0.209
AI460872	ADENOSYLMETHIONINE SYNTHETASE 2	1.689	0.362
AI460407	SRC1	1.897	0.329
AI522974	CATALASE	27.800	6.508
AI440706	GAMMA GLUTAMYLHYDROLASE	30.491	2.173
AI495553	PHOTOSYSTEM I REACTION CENTRE SUBUNIT IV PRECURSOR	72.244	11.696
AI461221	LIGHT-HARVESTING CHLOROPHYLL A/B BINDING PROTEIN	73.118	4.506
AI460590	ACC-OXIDASE	84.373	51.002
AI461105	PSAL GENE PRODUCT	109.713	43.222
AI461189	PHOSPHO-RIBULOKINASE PRECURSOR	137.482	11.247
AI441714	FERREDOXIN I PRECURSOR	250.799	124.644
AI460480	PHOTOSYSTEM I REACTION CENTRE SUBUNIT II PRECURSOR	280.076	90.559
AI495218	RIBULOSE BISP HOSPHATE CARBOXYLASE SMALL CHAIN 1 PRECURSOR	347.509	52.430
AI440658	PLASTOCYANIN PRECURSOR	395.870	192.265
AI495467	CHLOROPHYLL A-B BINDING PROTEIN OF LHCII TYPE I PRECURSOR	279.928	283.302
AI443426	TUBULIN ALPHA-2 CHAIN	87.528	8.683

Table 3. Highly abundant transcripts in both root and shoot tissue.

Gene ID	Description	Ave Ratio	SD
AI441931	COP9	0.741	0.179
AI461256	POLYUBIQUITIN PROTEIN	0.759	0.045
AI522836	ELONGATION FACTOR 1	0.962	0.141
AI440562	ASCORBATE PEROXIDASE	1.073	0.179
AI460407	SRC1	1.897	0.329
AI460591	70 KD HEATSHOCK PROTEIN	1.312	0.130
AI495462	ACTIN	1.162	0.209
AI495255	HISTONE H3.2, MINOR.	1.328	0.156
AI496483	ELONGATION FACTOR 1	1.430	0.189

respectively (data not shown). There were 95 sequences that were co-expressed at a moderate level in both the shoot and root (ratio 1.0) (data not shown). The majority of the expressed sequences were low abundance (<2,000 FSI). In the highly abundant class, as expected, many genes were well-characterised housekeeping or tissue-specific genes. Commonly expressed genes included actin and ubiquitin precursors, as well as other components of the intermediate metabolism.

Differential root : shoot gene expression

Dual labelling of cDNA probes with Cy3 and Cy5 fluorescent dyes allows simultaneous hybridisation of probes, which facilitates direct quantitative measurements of differential gene expression between two tissue types. The ratio of the two fluorescent signals can be used as a relative measure to determine the fold change in the differential expression of the sequence represented by the DNA spot on the microarray. Generally, differences greater than 2 fold are believed to be significant.

In this study we compared shoot tissue to root tissue (ratio = shoot/root), based on known physiological differences and function in these composite tissues. This comparison revealed many sequences that were differentially expressed (Fig. 1). The fold range of differential expression in the shoot was 3 fold to greater than 200 fold, with the majority around 5 fold. The fold range of differential expression in the root was 0.4 fold (equiv to 2.5 fold) to greater than 0.005 fold (equiv to 200 fold), with the majority around 0.04 fold (equiv to 25 fold). The degree of difference between the two tissue types tested reflects the level of physiological differences between the tissues, however the biological significance of the differences for all sequences is not known and interpretation depends on experimental reproducibility and biological sampling factors.

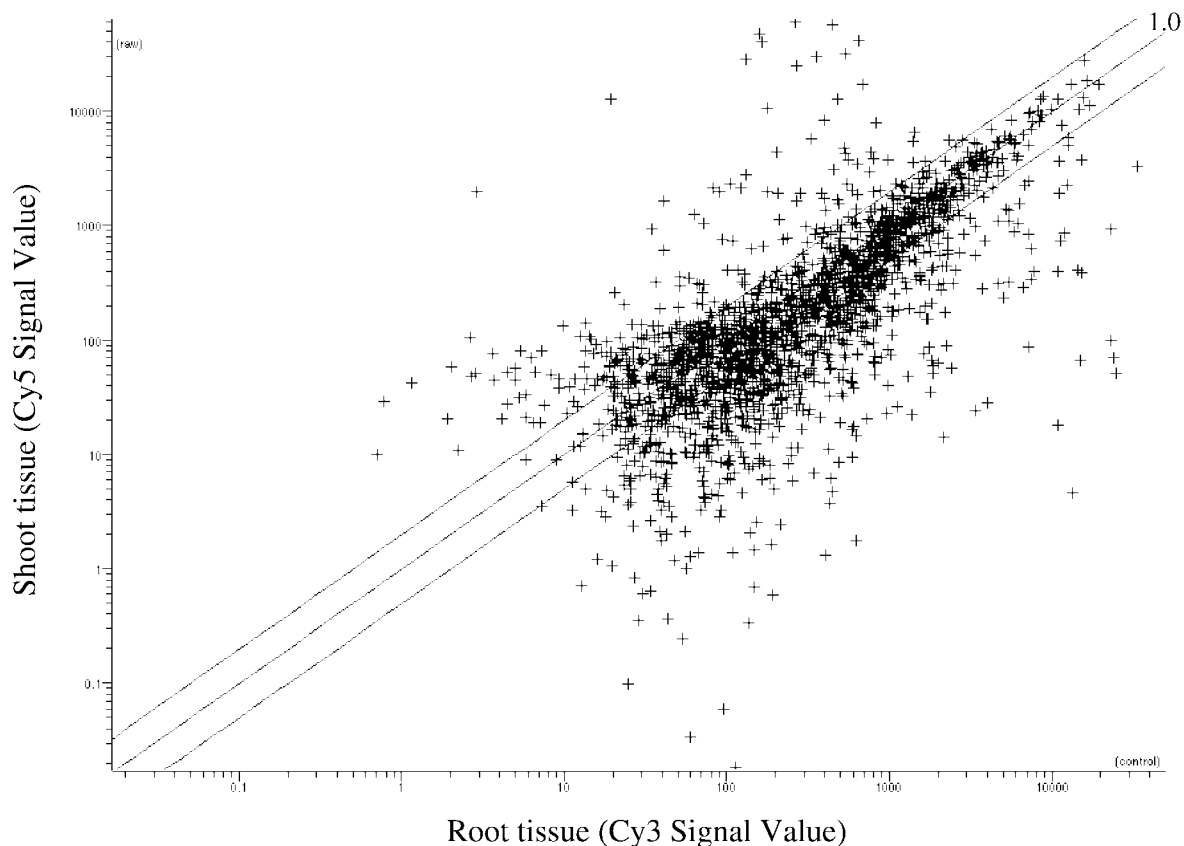


Figure 1. Log-scatter plot of signal values for all clones on the microarray. Signal intensities for each clone are shown. The diagonal lines represent normalised ratios, where the centre line represents a ratio of 1.0 (co-expression), with the outer lines representing 2-fold induction/repression.

The 50 most differentially expressed sequences in the root are shown in Table 4. The 47 most differentially expressed sequences in the shoot are shown in Table 5. The highest differentially expressed sequence in the root was a repetitive proline-rich cell wall protein 1 precursor (AI441027). The highest differentially expressed sequence in the shoot was a plastocyanin precursor (AI440658). In addition, 11 sequences show possible tissue-specific expression (>50 fold) in the shoot and 4 in the root.

The observed expression patterns coincide with expectations based on known physiological and biochemical functions in photosynthesis, nitrogen metabolism and cell wall architecture. Interestingly, the total number of differentially expressed ESTs from the shoot was similar to that of the root, despite using an array based on a root EST collection. In general one would expect a bias towards the printed tissue.

Not surprisingly, ESTs prominent in the photosynthetic apparatus (e.g., photosystem components, chloroplast proteins and light harvesting chlorophyll *a/b*-binding proteins) were differentially expressed in the shoot. Using other technologies Chang and Walling (1992) also found that chlorophyll *a/b*-binding protein accounted for 6.1% of total leaf mRNA, but

represented only 0.04% of root mRNA, suggesting a 150 fold differential expression.

Aquaporins were differentially expressed in the root tissue. Previous studies have also shown that aquaporins are predominantly expressed in roots (Miao and Verma 1993, Guenther and Roberts 2000), while calmodulin transcripts were rarely detected in roots (Lee et al. 1995). Interestingly, aquaporins have been shown to be modified by a calmodulin dependent Ca^{++} -protein kinase (Guenther and Roberts 2000), suggesting that low calmodulin levels suffice for biological function.

Further examples of similar patterns of differential gene expression can be demonstrated when comparing microarray data to previous studies using other technologies. For example, the small subunit of Rubisco (Meier et al. 1995, Silverthorne and Tobin 1990), extensin (Bucher et al. 1997), phenyl-propanoid pathway genes (Wingender et al. 1989), actin (Hightower and Meagher 1985), glycine-rich protein genes (De Oliveira et al. 1990), proline-rich protein genes (Wyatt et al. 1992) and auxin up or down-regulated genes (Datta et al. 1993). Several dormancy, Lea-type genes, as well as brassino-steroid protein genes were also shown to be differentially regulated, suggesting that shoot and root differ-

Table 4. Differential expression in root tissue.

Gene ID	Description	Ave Ratio	SD
AI441027	REPETITIVE PROLINE-RICH CELL WALL PROTEIN 1 PRECURSOR	0.003	0.001
AI441132	Novel	0.004	0.004
AI441396	Novel	0.008	0.004
AI461084	LEA5 PROTEIN	0.083	0.028
AI443989	RAD23 PROTEIN, ISOFORM II	0.021	0.019
AI444016	Novel	0.052	0.063
AI495517	IMMEDIATE-EARLY SALICYLATE-INDUCED GLUCOSYLTRANSFERASE	0.035	0.029
AI441105	REPETITIVE PROLINE-RICH CELL WALL PROTEIN 2 PRECURSOR	0.026	0.024
AI496590	HYPOTHETICAL PROTEIN	0.033	0.023
AI496573	REPETITIVEPROLINE-RICH CELL WALL PROTEIN 1 PRECURSOR	0.030	0.012
AI442523	FISSION YEAST	0.031	0.006
AI441001	METALLOTHIONEIN-LIKE PROTEIN 1	0.090	0.004
AI460497	Novel	0.052	0.028
AI442751	HYPOTHETICAL 35.7 KD PROTEIN	0.056	0.046
AI496350	PEROXIDASE PRECURSOR	0.039	0.028
AI496549	Novel	0.057	0.039
AI460620	HCAT-2A	0.044	0.024
AI440687	METALLOTHIONEIN-LIKE PROTEIN B	0.044	0.015
AI496538	EXTENSIN	0.049	0.018
AI495190	PEROXIDASE ATP EROX7A.	0.068	0.029
AI460757	SUCROSE SYNTHASE	0.082	0.036
AI460753	ASPARAGINE SYNTHETASE 2	0.065	0.036
AI441969	SALI3-2	0.074	0.019
AI494761	H(+)-TRANSPORTING ATPASE	0.076	0.038
AI441137	ASPARAGINE SYNTHETASE 2	0.078	0.028
AI460997	BETAINE-ALDEHYDEDE HYDROGENASE PRECURSOR	0.099	0.030
AI442735	ALPHA-AMYLASE/SUBTILISIN INHIBITOR PRECURSOR	0.145	0.087
AI495240	Novel	0.111	0.060
AI496205	ASPARAGINE SYNTHETASE 2	0.114	0.018
AI496671	SUCROSE SYNTHASE	0.114	0.085
AI496266	ANTIFREEZE PROTEIN	0.176	0.121
AI441549	Novel	0.122	0.063
AI496438	PKF1 PROTEIN	0.131	0.099
AI440678	Novel	0.132	0.052
AI441983	EARLY NODULIN 36A (ENOD40)	0.167	0.054
AI495554	Novel	0.168	0.050
AI441704	BLUE COPPER-BINDING PROTEIN II	0.303	0.291
AI440686	MIPC	0.187	0.030
AI496108	PEROXIDASE PRECURSOR	0.190	0.046
AI507717	Novel	0.205	0.052
AI441335	HYPOTHETICAL PROTEIN	0.216	0.074
AI496129	Novel	0.217	0.029
AI442288	ENDOCHITINASE PRECURSOR, BASIC	0.233	0.083
AI496156	EXTENSIN	0.263	0.029
AI460606	Novel	0.270	0.032
AI496396	DORMANCY-ASSOCIATED PROTEIN	0.273	0.062
AI443919	60S RIBOSOMAL PROTEIN L36	0.286	0.047
AI461078	PROLINE-RICH 14 KDA PROTEIN	0.290	0.132
AI522819	WATER-STRESS INDUCED TONOPLAST INTRINSIC PROTEIN	0.355	0.034
AI461005	Novel	0.397	0.049
AI496329	HYPOTHETICAL 17.7 KD PROTEIN	0.436	0.104
AI461236	PUTATIVE WATER CHANNEL PROTEIN	0.451	0.078

Table 5. Differential expression in shoot tissue.

Gene ID	Description	Ave Ratio	SD
AI460489	BETA-FRUCTOFURANOSIDASE PRECURSOR	3.212	1.167
AI496533	PUTATIVE CINNAMOYL-CO-A REDUCTASE	18.310	6.828
AI496426	MYO-INOSITOL-1-PHOSPHATE SYNTHASE	3.806	0.998
AI461228	BRASSINOSTEROID-REGULATED PROTEIN BRU1	4.165	0.934
AI441739	Novel	4.368	1.618
AI441395	INDOLE-3-ACETIC ACID INDUCED PROTEIN ARG7	4.384	1.016
AI461233	PECTIN ACETYL-ESTERASE PRECURSOR	4.428	1.244
AI494760	DIHYDROLIPOAMIDE DEHYDROGENASE PRECURSOR	4.459	0.577
AI495099	CHLOROPLAST 50S RIBOSOMAL PROTEIN L1 PRECURSOR	4.848	3.764
AI495256	KETOL-ACID-REDUCTOISOMERASE PRECURSOR	5.072	1.074
AI495130	EXPANSIN	5.089	2.410
AI495592	CCA1	5.200	0.849
AI522857	CAPI PROTEIN	5.340	1.240
AI495572	37 KD CHLOROPLAST INNER ENVELOPE MEMBRANE PROTEIN PRECURSOR	5.798	1.345
AI460668	HYPOTHETICAL 23.5 KD PROTEIN	6.757	1.550
AI495575	SALT-TOLERANCE PROTEIN	6.796	3.154
AI494876	L-ASPARAGINASE	7.307	2.386
AI441635	CHLOROPLAST 31 KD RIBONUCLEOPROTEIN PRECURSOR	7.722	2.195
AI460461	THIOREDOXIN	19.966	1.781
AI522863	REMORIN	8.083	2.425
AI507878	F4P9.33 PROTEIN	8.376	5.438
AI442696	GAG PROTEINS	9.023	1.009
AI495564	Novel	10.565	4.233
AI441990	NUCLEOSIDE-TRIPHOSPHATASE	11.405	2.023
AI460688	ADR11 PROTEIN (Auxin)	11.939	2.049
AI507755	Novel	12.539	2.968
AI442189	Novel	22.618	6.992
AI496052	Novel	16.056	13.334
AI460509	HYPOTHETICAL 5.5KD PROTEIN YCF17	19.288	4.510
AI460786	PHOTOSYSTEM II 22KD PROTEIN PRECURSOR	33.374	4.029
AI460444	ALANINE AMINOTRANSFERASE	23.939	3.787
AI522974	CATALASE	27.800	6.508
AI443769	BAS1 PROTEIN	28.073	10.927
AI440706	GAMMA GLUTAMYLHYDROLASE	30.491	2.173
AI441371	AMINO ACID TRANSPORTER	34.899	8.987
AI495863	Novel	36.363	25.523
AI495553	PHOTOSYSTEM I REACTION CENTRE SUBUNIT IV PRECURSOR	72.244	11.696
AI461221	LIGHT-HARVESTING CHLOROPHYLL A/B BINDING PROTEIN	73.118	4.506
AI460590	ACC-OXIDASE	84.373	51.002
AI461105	PSAL GENE PRODUCT	109.713	43.222
AI461189	PHOSPHORIBULOKINASE PRECURSOR	137.482	11.247
AI441714	FERREDOXIN I PRECURSOR	250.799	124.644
AI460480	PHOTOSYSTEM I REACTION CENTRE SUBUNIT II PRECURSOR	280.076	90.559
AI495218	RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL CHAIN 1 PRECURSOR	347.509	52.430
AI440658	PLASTOCYANIN PRECURSOR	395.870	192.265
AI495467	CHLOROPHYLL A-B BINDING PROTEIN OF LHCII TYPE I PRECURSOR	279.928	283.302
AI443426	TUBULIN ALPHA-2 CHAIN	87.528	8.683

entiation involves a suite of metabolic as well as regulatory pathways.

Of special interest was the differential expression of ACC oxidase in the shoot tissue. The plants used in this study were grown under optimal conditions (Conviron growth cabinet) and showed no stress or disease symptoms; yet ACC oxi-

dase, an enzyme critical in the biosynthesis of the plant stress hormone ethylene, was differentially expressed (84 fold) in the shoot tissue. It is yet undetermined whether the increased mRNA level is reflected in equally elevated enzyme activity and whether the differential is part of a normal physiological balance or a response to an unknown and subtle stress.

SALI3-2, an aluminium induced gene, was also differentially expressed in roots (13 fold) suggesting some form of physiological stress perhaps derived from the sand-vermiculite mix. Detection of such unexpected responses further demonstrates the physiological complexity of plants and emphasises the need for holistic and integrative methods such as microarray expression analysis to recognise gene regulatory networks instead of «mere» pathways.

Overall, the reliability, dynamic range (expression ratios over 5–6 orders of magnitude) and broad representation of cDNA clones allows us to formulate patterns of coordinate gene expression, not only within the same tissue, but also between tissue types.

Expression of novel coding sequences

Of the 4,089 Unigene EST sequences used in this study, there were 1,603 novel sequences that have no significant similarities to any known cDNA sequences in current public databases. cDNA microarrays provide an efficient approach to assess the function of these sequences by observing their expression patterns. There were 3 novel sequences that were highly expressed (>10,000 FSI) in the root, and 15 novel sequences that were differentially expressed in roots. In contrast, there were only 6 novel sequences that were differentially expressed in the shoots. It is possible that there was some bias towards novel sequences that are differentially expressed in the roots, since the EST library was constructed from 8-day-old total root tissue. In the moderate range of gene expression (between 2,000 and 10,000 FSI) there were 10 novel sequences that were co-expressed in the root and shoot tissue (ratio 1.0).

These data highlight the utility of cDNA microarrays for identifying novel genes in soybean, in contrast to past approaches where only a few novel genes have been identified at a time. It should also be noted that other cDNA microarray

approaches have revealed novel plant sequences involved in nutrient responses, seed formation, anther/pistil differentiation and circadian/diurnal regulation (Wang et al. 2000, Girke et al. 2000, Endo et al. 2002, Schaffer et al. 2001), further validating this approach.

Verification of microarray data using Northern blot analysis and quantitative real-time RT-PCR

To further confirm the expression patterns of microarray data 10 sequences were selected for verification (Table 6) using Northern blot analysis and quantitative real-time RT-PCR. The source of RNA used for verification studies was a biological replicate, plants were grown at a different time under identical conditions, pooled and used for RNA isolation. The selected sequences were from different abundance classes and cellular processes. Figure 2 shows the transcript levels of the 10 genes analysed, methylene blue was used as a control for equal RNA loading. In 9 out of 10 cases the Northern blot analysis strongly supports the microarray data. Sequences 1, 2, 5, 6 and 10 show strong differential expression in the shoot tissue and 3, 7, 8 and 9 show strong differential expression in the root tissue. While sequence number 4 (tubulin) shows co-expression, in contrast to the microarray data which shows differential expression in the shoot tissue. Table 7 shows the approximate transcript sizes of the 10 sequences which range from 1 kb to 2.8 kb.

Figure 3 illustrates a typical plot of data generated by real-time RT-PCR using SYBR green detection. The PCR signal is initially below the level of detection and increases with cycle number until it crosses a threshold. The threshold was set to the midlinear portion of the log ΔR_n versus cycle plot. The cycle number at which the signal crosses the threshold is defined as C_t . If a cDNA sample has more copies of a template then that sample will cross the threshold at an earlier cycle number, compared to one containing fewer copies of the

Table 6. Ten sequences were selected for verification of gene expression using Northern blot analysis and quantitative real-time RT-PCR. The ratios determined by microarray and quantitative real-time RT-PCR data are shown.

Gene Number	Gene ID	Putative function	Ave Ratio Microarray	SD Microarray	Ratio RT-PCR
1	AI461105	PSAL GENE PRODUCT	109.713	43.222	149.086
2	AI460480	PHOTOSYSTEM I REACTION CENTRE SUBUNIT II PRECURSOR	280.076	90.559	1221.976
3	AI440686	MIPC, aquaporin	0.187	0.030	0.099
4	AI443426	TUBULIN ALPHA2 CHAIN	87.528	8.683	1.094
5	AI460590	ACC-OXIDASE	84.373	51.002	218.275
6	AI460688	ADR11 PROTEIN	11.940	2.049	15.137
7	AI441137	ASPARAGINE SYNTHETASE 2	0.078	0.028	0.084
8	AI441027	REPETITIVE PROLINE-RICH CELL WALL PROTEIN 1 PRECURSOR	0.004	0.001	0.001
9	AI441969	SALI3-2, aluminium induced	0.074	0.019	0.112
10	AI522974	CATALASE	27.800	6.508	46.046
Control	AI507761	ACTIN 2/7	0.832	0.099	0.959

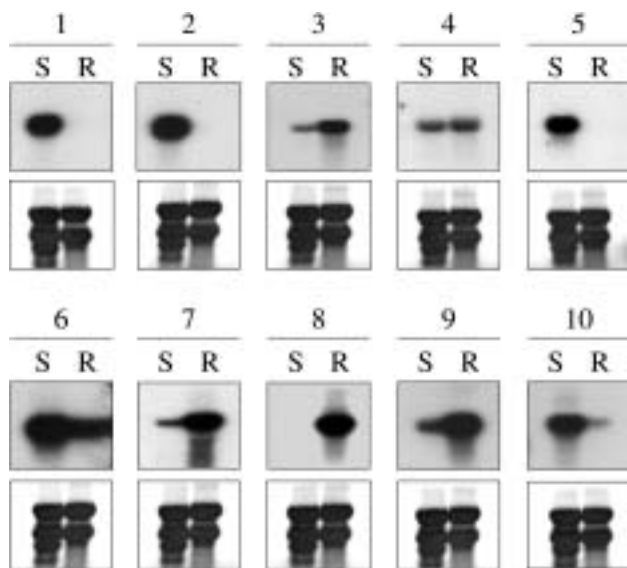


Figure 2. Northern blot results. Expression patterns detected by Northern blot analysis in different tissue types of soybean (S, shoot; R, root) were compared. The accession number for each gene is shown in Table 6. The ribosomal RNA bands were also included as sample-loading standards for Northern blot analysis.

template. Since SYBR green I indiscriminately binds to double-stranded DNA, other products in the PCR reaction such as primer-dimers or contaminating genomic DNA may also be detected along with the target gene (Morrison et al. 1998). To verify that the SYBR green dye detected only one PCR product, the samples were subjected to a heat dissociation protocol following the final cycle of the PCR. Heat dissociation of oligonucleotides detects differences in melting temperature and will produce a single dissociation peak for each oligonu-

cleotide within a 2 °C difference in melting temperature (Ririe et al. 1997). A typical example is shown in Figure 4, demonstrating the presence of only one product in the reaction. The flat lines underneath the peak represent the minus reverse transcriptase control to test for genomic DNA contamination. Duplicate cDNAs from each tissue type were amplified using primers designed for all 10 sequences and were detected using SYBR green. A control gene (actin) was also included. The amount of fluorescent product was transformed equal to $2^{-\Delta C_t}$ in order to convert C_t (logarithmic value) to a linear value. The data are presented in Table 8 as the fold ratio, and were not normalised to the internal control gene (actin) because the threshold value for actin was identical in both tissue types. In 9 out of 10 cases the quantitative real-time RT-PCR data strongly supports the microarray data. Sequence number 4 (tubulin) shows co-expression using real-time RT-PCR, in contrast to the microarray data which shows differential expression in the shoot tissue. This is in strong agreement with the Northern blot analysis (Fig. 2).

To resolve this inconsistency, we conducted a repeat hybridisation using new microarrays from a different print-run and arrayer. We found that the new batch of microarrays showed a ratio of approximately 1.0 for sequence number 4 (tubulin), while the other 9 sequences had similar ratios as shown previously (data not shown). This suggests that a printing artefact caused the inconsistency for sequence number 4. Thus, the importance of verifying the expression pattern of genes of interest, especially if they belong to multigene families. The log ratios for real-time RT-PCR data and microarray data were plotted (Fig. 5) and it was found that there was a strong linear relationship with a r^2 value of 0.99 (excluding sequence 4). This indicates that there is a strong linear relationship between real-time RT-PCR data and microarray data over a 5–6 log dynamic range using SYBR green detection.

Table 7. Approximate transcript sizes of the 10 selected sequences, determined by Northern blot analysis.

Gene Number	Gene ID	Putative function	Approx. transcript size (bases)
1	AI461105	PSAL GENE PRODUCT	1.000
2	AI460480	PHOTOSYSTEMI REACTION CENTRE SUBUNIT II PRECURSOR	1.100
3	AI440686	MIPC. aquaporin	1.700
4	AI443426	TUBULIN ALPHA2 CHAIN	2.500
5	AI460590	ACC-OXIDASE	1.500
6	AI460688	ADR11PROTEIN	1.500
7	AI441137	ASPARAGINE SYNTHETASE 2	2.800
8	AI441027	REPETITIVE PROLINE-RICH CELL WALL PROTEIN 1 PRECURSOR	1.700
9	AI441969	SAL13-2. aluminium induced	1.600
10	AI522974	CATALASE	2.800

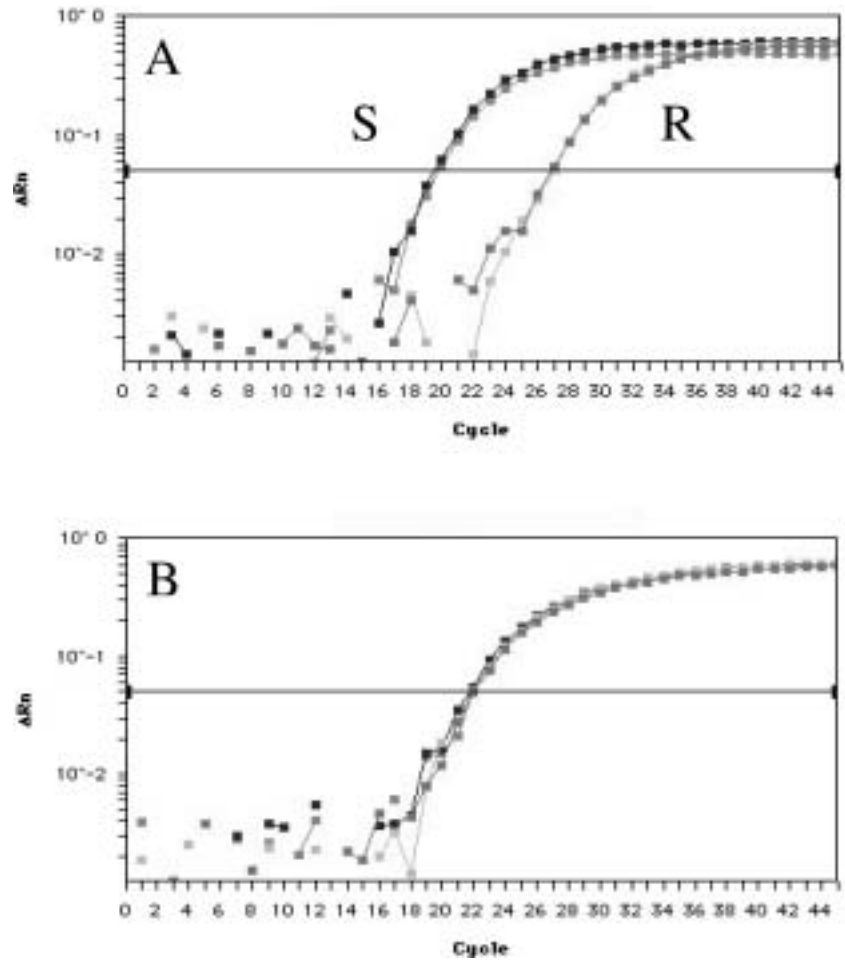


Figure 3. Real-time RT-PCR analysis in different tissue types of soybean (S, shoot; R, root) were compared. (A) Typical plot of the ΔR_n versus cycle number obtained with SYBR green detection for differentially expressed genes. Duplicate samples of cDNA from root and shoot tissue are shown. The line represents the threshold arbitrarily set in the middle of the log ΔR_n plot. (B) Plot of the ΔR_n versus cycle number obtained with SYBR green detection for a co-expressed gene.

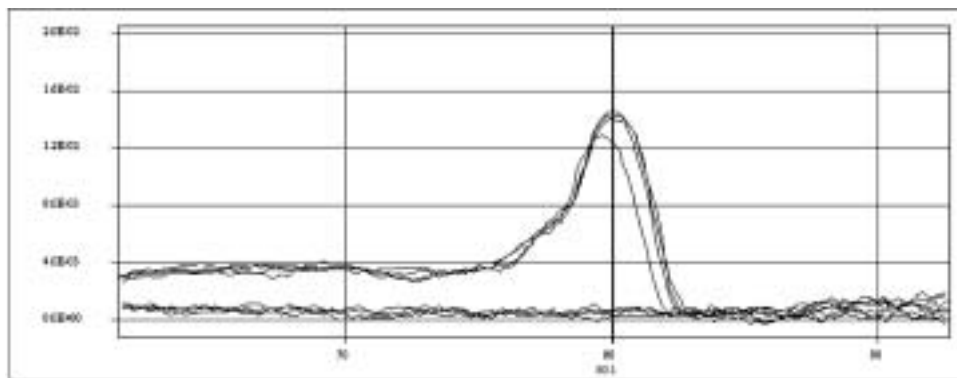


Figure 4. Heat dissociation protocol. All samples were amplified by 45 cycles of PCR using primers designed for each of the 10 selected sequences using SYBR green detection. Following the final PCR cycle the samples were subjected to a heat dissociation protocol over the indicated temperature range. The lines under the peak on the graph represent the minus reverse transcriptase controls.

Table 8. Quantitative real-time RT-PCR results for 10 selected sequences are shown for each tissue type of soybean (R = root, S = shoot). Ct = cycle threshold number, $\Delta\Delta Ct$ = difference in cycle number between root and shoot tissue respectively, ratio calculated as $2^{-\Delta\Delta Ct}$.

Gene Number	R1 Ct	R2 Ct	R Ave Ct	S1 Ct	S2 Ct	S Ave Ct	$\Delta\Delta Ct$ R-S	Ratio
1	26.790	26.950	26.870	19.510	19.790	19.650	-7.220	149.086
2	29.400	31.020	30.210	19.930	19.980	19.955	-10.255	1221.976
3	19.530	19.620	19.575	22.690	23.120	22.905	3.330	0.099
4	22.000	22.040	22.020	21.790	21.990	21.890	-0.130	1.094
5	30.220	30.310	30.265	22.240	22.750	22.495	-7.770	218.275
6	24.920	25.030	24.975	21.020	21.090	21.055	-3.920	15.137
7	20.220	20.640	20.430	23.910	24.110	24.010	3.580	0.084
8	20.540	21.540	21.040	31.400	31.580	31.490	10.450	0.001
9	20.080	20.150	20.115	23.140	23.400	23.270	3.155	0.112
10	27.760	27.910	27.835	22.190	22.430	22.310	-5.525	46.046
Control (1)	23.210	23.270		22.580	23.590			
Control (2)	23.350	23.490	23.330	23.660	23.730	23.390	0.060	0.959

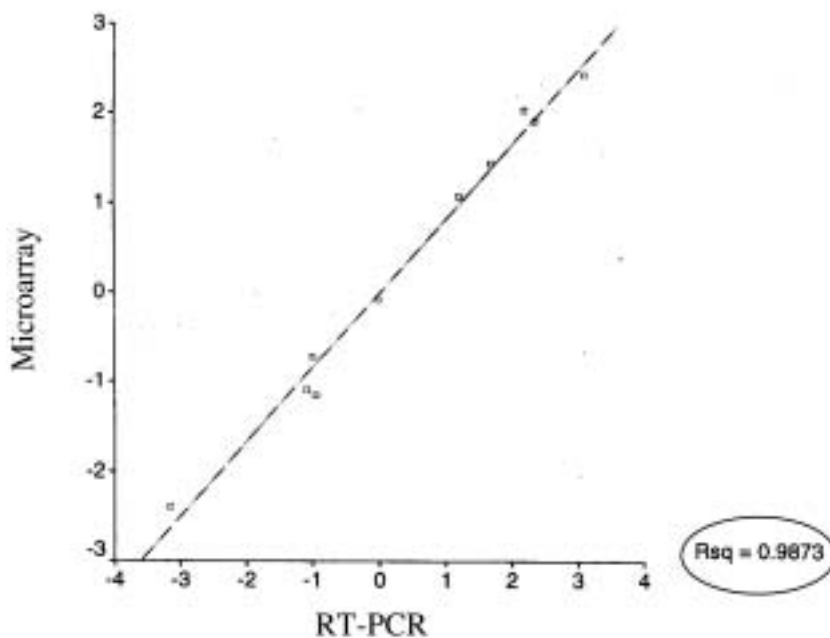


Figure 5. Relationship between quantitative real-time RT-PCR and microarray data. Plot of selected genes log ratios determined using quantitative real-time RT-PCR and microarrays. Gene accession numbers and ratios are shown in Table 6. The dynamic range for each method is presented. The r^2 was determined using least-squares linear regression method.

Conclusions

This paper opens the possibility to investigate biological processes specific to legumes. Our research group is especially interested in the processes of nodulation, nitrogen fixation and interactions with mycorrhizae, all of which are not possible in the model plant *Arabidopsis*. Legumes can also provide biological information on novel biosynthetic pathways including novel saponins and derivatives from the phenyl-propanoid pathway, known to be critical for plant microbe interactions, human nutrition and plant stress physiology (e.g., flavonoids, isoflavones, antioxidants, lipids). Some of these substan-

ces such as daidzein and genistein are known phytoestrogens; some are anti-tumorigenesis factors presumably acting through an activation of cytochrome p450 mono-oxygenases.

This paper emphasised the analysis of known genes for two purposes (a) to use previous knowledge to validate the microarray data obtained and (b) to demonstrate the potential of parallel gene expression analysis approaches in soybean and eventually whole genome scale expression approaches. The accuracy of microarray data was validated by reproducibility of the microarray results and correlation with data from other technologies, such as Northern blot analysis and quantitative real-time RT-PCR. We also demonstrate that with good

quality control in the isolation of the initial RNA, and quality assurance during array construction and hybridisation, primary data can be used to discover coordinate gene expression patterns.

Identification of genes expressed in a specific tissue type can provide valuable insights into basic molecular processes. We identified both constitutive and differentially expressed sequences in soybean tissues. Identification of constitutive and tissue-specific genes is important to isolate constitutive or organ/tissue specific promoters. The promoter regions of co-regulated genes can be compared to identify common regulatory elements. Using this approach microarrays may facilitate the dissection of regulatory pathways and provide opportunities to genetically engineer metabolic pathways. Identification of co-regulated genes is the first step towards understanding the regulation of gene expression networks and assigning function to new genes. With an increasing database of soybean ESTs the functional characterisation of novel sequences becomes the challenge for soybean genomics. Using microarrays, mutant lines or treated biological samples (and their controls) can be transcriptionally profiled. By comparing alterations of novel gene expression profiles a potential function can be assigned. This can be further confirmed by reverse genetics approaches.

Acknowledgements. We thank Mr Bob Simpson at the Real-Time PCR Facility in the Department of Biochemistry and Molecular Biology, The University of Queensland, for his assistance and access. Dr Maguire is supported by a University of Queensland Post Doctoral Research Fellowship and Dr Grimmond is supported by a NHMRC Career Development Award. Microarray infrastructure support was provided in part by the ARC Special Research Centre for Functional and Applied Genomics. The Public EST Project for Soybean is supported by the North Central Soybean Research Board and the United Soybean Board through a Specific Cooperative Agreement No. 58-3625-8-117 through the USDA-Agricultural Research Service.

References

- Aharoni A, Keizer LCP, Bouwmeester HJ, Sun Z, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen AMML, DeVos RCH, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ, O'Connell AP (2000) Identification of the *SAAT* gene involved in strawberry flavour biogenesis by the use of DNA microarrays. *Plant Cell* 12: 647–661
- Arabidopsis Genome Initiative (2001) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 409: 796–815
- Bucher M, Schroeer B, Willmitzer L, Riesmeier JW (1997) Two genes encoding extensin-like proteins are predominantly expressed in tomato root hair cells. *Plant Mol Biol* 35: 497–508
- Chang YC, Walling LL (1992) Chlorophyll *a/b*-binding protein genes are differentially expressed during soybean development. *Plant Mol Biol* 19: 217–230
- Datta N, LaFayette PR, Kroner PA, Nagao RT, Key JL (1993) Isolation and characterisation of three families of auxin down-regulated cDNA clones. *Plant Mol Biol* 21: 859–869
- De Oliveira DE, Seurinck J, Inze D, Van Montagu M, Botterman J (1990) Differential expression of five *Arabidopsis* genes encoding glycine-rich proteins. *Plant Cell* 2: 427–436
- Endo M, Matsubara H, Kokubun T, Masuko H, Takahata Y, Tsuchiya T, Fukuda H, Demura T, 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
- Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J (2000) Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol* 124: 1570–1581
- Guenther JF, Roberts DM (2000) Water-selective and multifunctional aquaporins from *Lotus japonicus* nodules. *Planta* 210: 741–748
- Hedge P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J (2000) A concise guide to cDNA microarray analysis. *BioTechniques* 29: 548–562
- Herridge DF (1977) Carbon and nitrogen nutrition of two annual legumes. PhD Thesis. University of Western Australia, Perth, Australia
- Hightower RC, Meagher RB (1985) Divergence and differential expression of soybean actin genes. *EMBO J* 4: 1–8
- Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ (2001) Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell* 13: 889–905
- Lee SH, Kim JC, Lee MS, Heo WD, Seo HY, Yoon HW, Hong JC, Lee SY, Bahk JD, Hwang I (1995) Identification of a novel divergent calmodulin isoform from soybean which has differential ability to activate calmodulin-dependent enzymes. *J Biol Chem* 270: 21806–21812
- Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL, Dietrich RA (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nature Genet* 26: 403–410
- Meier I, Callan KL, Fleming AJ, Grissem W (1995) Organ-specific differential regulation of a promoter subfamily for the ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit genes in tomato. *Plant Physiol* 107: 1105–1118
- Miao GH, Verma DP (1993) Soybean nodulin-26 gene encoding a channel protein is expressed only in the infected cells of nodules and is regulated differently in roots of homologous and heterologous plants. *Plant Cell* 5: 781–794
- Morrison TB, Weis JJ, Wittwer CT (1998) Quantification of low-copy transcripts by continuous SYBR green I monitoring during amplification. *BioTechniques* 24: 954–962
- Reymond P, Weber H, Damond M, Farmer EE (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell* 12: 707–719
- Ririe KM, Rasmussen RP, Wittwer CT (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem* 245: 154–160
- Ruan Y, Gilmore J, Conner T (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J* 15: 821–833
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A laboratory manual*. 2nd Ed. Cold Spring Harbour Laboratory Press, USA
- Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E (2001) Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant Cell* 13: 113–123

- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with complimentary DNA microarray. *Science* 270: 467–470
- Schenk PN, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Coordinated plant defence responses in *Arabidopsis* revealed by microarray analysis. *Proc Natl Acad Sci USA* 97: 11655–11660
- Seki M, Narrusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carnici P, Hayashizaki Y, Shinozaki K (2001) Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell* 13: 61–72
- Shoemaker R, Keim P, Vodkin L, Erpelding J, Coryell V, Khanna A, Bolla B, Marra M, Hillier L, Kucaba T, Martin J, Beck C, Wylie T, Underwood K, Steptoe M, Theising B, Allen M, Bowers Y, Person B, Swaller T, Gibbons M, Pape D, Harvey N, Schurk R, Ritter E, Kohn S, Shin T, Jackson Y, Cardenas M, McCann R, Waterston R, Wilson R (1999) Public Soybean EST Project (unpublished)
- Silverthorne J, Tobin EM (1990) Post-transcriptional regulation of organ-specific expression of individual rbcS mRNAs in *Lemna gibba*. *Plant Cell* 2: 1181–1190
- Wang R, Guegler K, LaBrie ST, Crawford NM (2000) Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell* 12: 1491–1509
- Wingender R, Rohrig H, Horicke C, Wing D, Schell J (1989) Differential regulation of soybean chalcone synthase genes in plant defence, symbiosis and upon environmental stimuli. *Mol Gen Genet* 218: 315–322
- Wyatt RE, Nagao RT, Key JL (1992) Patterns of soybean proline-rich protein gene expression. *Plant Cell* 4: 99–110
- Yu J, Nickels R, McIntosh L (2001) A genome approach to mitochondrial-nuclear communication in *Arabidopsis*. *Plant Physiol Biochem* 39: 345–353
- Zhu T, Budworth P, Han B, Brown D, Chang HS, Zou G, Wang X (2001) Toward elucidating the global gene expression patterns of developing *Arabidopsis*: parallel analysis of 8300 genes by a high-density oligonucleotide probe array. *Plant Physiol Biochem* 39: 221–242