January, August & Final Reports

REPORTS

Part 1 - Summary Details

Please use your TAB key to complete part 1 & 2.

CRDC Project Number: **CSP119C**

January Report:	☐	Due 29-Jan-01
August Report:	☐	Due 03-Aug-01
Final Report:	☒	Due within 3 months of project completion

Project Title: Use of miroarrays to study gene expression and to identify genes involved in cotton fibre initiation and development

Project Commencement Date: 1/7/2000

Project Completion Date: 30/6/2001

Research Program: Plant Breeding and Biotechnology

Part 2 – Contact Details

Administrator: Dianne Rosson
Organisation: CSIRO Plant Industry
Postal Address: GPO Box 1600 Canberra ACT 2601
Ph: 02-62465277 **Fx:** 02-62465000 **E-mail:** dianne.rosson@csiro.au

Principal Researcher: Yingru Wu
Organisation: CSIRO Plant Industry
Postal Address: GPO Box 1600 Canberra ACT 2601
Ph: 02-62464914 **Fx:** 02-62465000 **E-mail:** yingru.wu@csiro.au

Supervisor: Danny Llewellyn
Organisation: CSIRO Plant Industry
Postal Address: GPO Box 1600 Canberra ACT 2601
Ph: 02-62465470 **Fx:** 02-62465000 **E-mail:** danny.llewellyn@csiro.au

Researcher 2 (Name & position of additional researcher or supervisor).
Organisation:
Postal Address:
Ph: **Fx:** **E-mail:**

Signature of Research Provider Representative:

Part 3 – Final Report Format

1. Background

Cotton provides about 55% of the fibre in textile manufacturing globally and is an important contributor to the Australian economy. The cotton fibre is a single cell hair (trichome) growing out from the surface of the cotton seed epidermis. Little is known about the molecular basis for the control of which cells initiate fibre growth and how that growth is regulated to produce fibres of a desired quality. We have initiated a project to study the genes expressed during the early stages of fibre cell initiation in cotton using both directed approaches to target genes that might be expected to have a role in fibre initiation, as well as a genomics approach to look for novel genes expressed in the early stages of fibre growth. cDNA libraries from early stage cotton ovules have been produced and we have started sequencing clones at random from these libraries (supplemented by a one year grant from CRDC - CSP110C) and we will use bioinformatics techniques (computer database searching) to try to identify clones that might have a role in the control of fibre development processes (particularly genes known to control other genes – transcription factor genes). This can generate thousands of candidates and the next stage of the process is to narrow these genes down to a workable number based on their patterns of expression in plants. To do this we can use micro-arrays – these are glass slides onto which small amounts of each of many thousands of the genes are spotted in an ordered sequence. By then using fluorescently labelled probes of the total RNA from particular stages or conditions we can use a dedicated scanner and computer software to measure the level of expression of all 10,000 or more genes under different conditions, such as at different stages of fibre initiation and development or between cultivars with different quality attributes. Genes whose expression goes up, down or stays the same can be identified and by doing different comparisons we can hope to identify those genes controlling the initiation process or quality traits.

2. List the project objectives and the extent to which these have been achieved.

July 2000 – July 2001: To amplify 10,000 cDNA clones expressed in cotton ovules and array them on glass slides. These clones will mostly be from CSIROs own fibre cDNA libraries but may include clones from overseas laboratories if these can be obtained. The slides will be used to explore the expression patterns of the genes at different stages of development and in comparisons between a normal and lintless mutant of cotton or between cotton lines with different quality traits.

10,410 cDNA clones from our cotton ovule libraries have been amplified and printed onto 96 microarray slides. The slides have been used to study the gene expression pattern of a lintless cotton mutant as compared to a wild type cotton. A statistical package for microarray data analysis tRMA was used to extract lists of differentially expressed genes. Sequencing of these differentially expressed genes and further microarray studies of gene expression patterns at different developmental stages are in progress.

3. How has your research addressed the Corporations three outputs: Sustainability, profitability and international competitiveness, and/or people and community?

The last few years have seen rapid advances in the analysis of plant genomes that have increased our knowledge of gene structure and function in plants and shown potential for the identification and manipulation of important genes for the agronomic improvement of crop plants through biotechnology. Multi-national companies (such as Monsanto, Dow and Novartis) are generating genome sequences and libraries of expressed genes from agricultural crops from which they hope to isolate key target genes specifying agronomic traits or protection of plants from pests and diseases. These genome resources will be propriety and protected by patents so will not be available for Australian researchers or Industry except through license arrangements that are generally not to the farmer's advantage. Australian researchers must be active in these areas if we are to participate at an equal level with the multi-nationals over the next decade. CSIRO Plant Industry has initiated genomics approaches to the analysis of genes in cereals such as wheat, barley and rice (through CSIRO and Graingene), but the importance of cotton to the Australian economy and our demonstrated capacity to use biotechnological approaches for cotton improvement warrant a concerted effort in cotton with a focus on key traits of value to the industry, such as fibre quality genes, and disease and pest tolerance genes. The process of gene discovery will generate new Intellectual Property for Australia. Ownership of this IP will enable us to use it in our variety development program and we will be able to trade access under suitable conditions for return access to other proprietary traits of value for Australian cotton. A better understanding of the molecular processes in fibre initiation and development will allow us to design new strategies to genetically improve fibre yield and quality in Australian varieties.

4. Detail the methodology and justify the methodology used.

Microarrays are formed by robotically depositing specific fragments of DNA at indexed locations on microscope slides. The DNA fragments can originate from a variety of sources including anonymous cDNA clones, EST clones, anonymous genomic clones, or DNA amplified from open reading frames (ORFs) found in sequenced genomes.

Once produced, the microarrays are hybridised with fluorescently-labeled mRNA-derived probes and the hybridised probes on the array are excited by light. The fluorescent signal emitted from each spot is a reflection of the abundance of the corresponding sequence in the original probe. Microarray technology is ideally suited for making pair-wise comparisons of samples. Two fluorescent tags, with different excitation and emission optima, can be used to label two distinct probes (eg. two mRNA populations from physiologically or genetically distinct samples). The two probes are mixed and allowed to hybridise to the same microarray. For each spot on the microarray, the ratio of fluorescence emission at the two wavelengths reflects the ratio of the abundance of that mRNA species in the two probes.

Microarray is one of the most powerful tools that have recently been developed to bridge the gap between sequence information and functional genomics. The power lies on its scale, sensitivity and quantitative nature of data output. Theoretically, gene expression patterns of a whole plant genome (20,000-30,000 genes) can be studied simultaneously. Because the detection is fluorescence based, the signal output is very sensitive, and individual mRNA species can be detected at a threshold of 1 part in 100 000 to 1 part in 500 000. Because the output is quantitative, subtle changes in gene expression can be detected, in addition to the more dramatic changes observed with such techniques as subtractive hybridisation and differential display. The application of this powerful technology to cotton fibre genomics was only made possible by the installation at CSIRO of a robotic microarrayer and scanners purchased from GrainGene funds and additional operating funds from the CRDC for

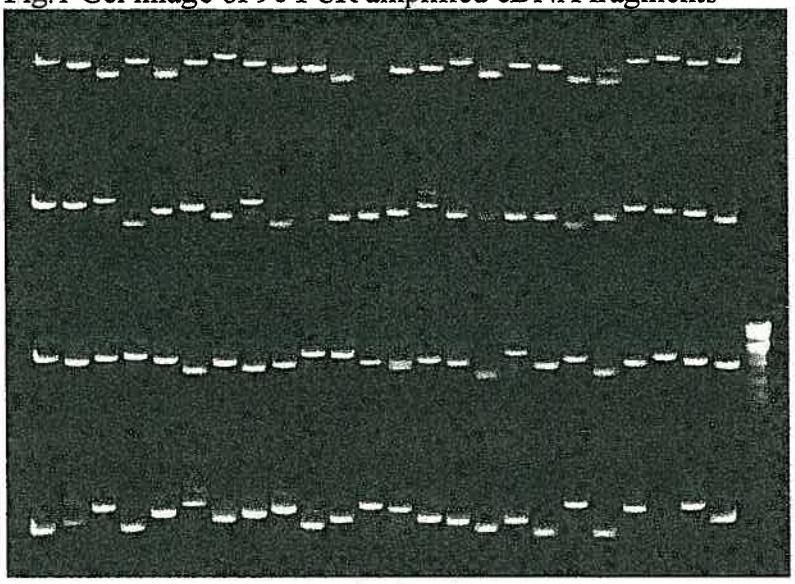
microarray consumables and some additional minor equipment need to be used in these microarray studies in cotton.

5. Detail results including the statistical analysis of results.

a. Microarray slides generation

A total of 10,410 cDNA clones from our cotton ovule cDNA libraries were amplified by PCR. The PCR reactions were carried out in 96 well PCR plates and the amplified products were verified on gels. Figure 1 shows the typical PCR products from one PCR plate.

Fig.1 Gel image of 96 PCR amplified cDNA fragments



The 10,410 cDNA clones are composed of ~3000 sequenced cDNA clones (CSP110C) and ~7000 unsequenced cDNA clones from our libraries, plus some other known cotton genes from our collaborators at Plant Industry. Table 1 lists the details of these sources.

Table 1. Clones amplified and arrayed on the fibre microarray

Source cDNA Libraries	-3 to -1 dpa* Ovule Library	0 dpa Ovule Library	Normalised 0dpa Ovule Library
Sequenced Clones	1000	1000	1000
Unsequenced Clones	4000	0	3000
Total	5000	1000	4000

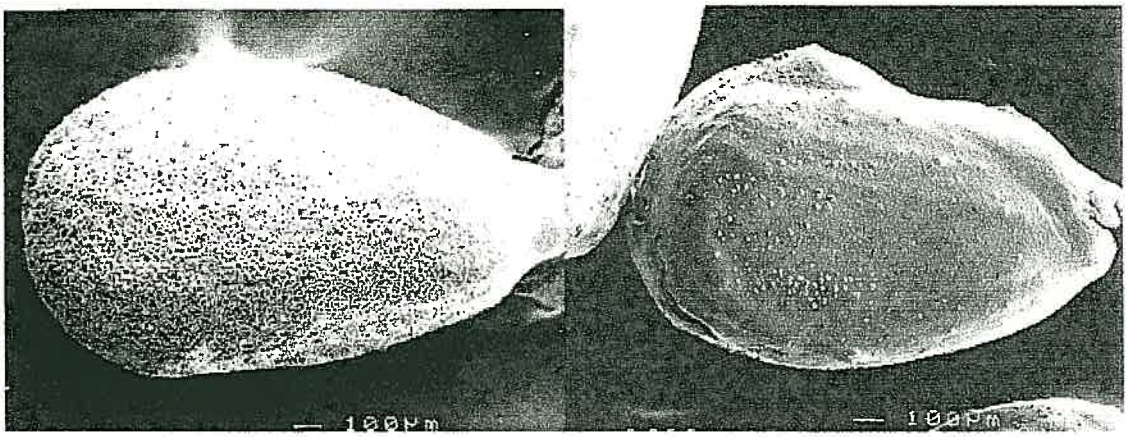
* dpa: days post anthesis

The amplified cDNA clones were purified and transferred to 384 plates used for robotic printing of microarray slides. The first bach of 96 slides were printed with 24 pins which generated 24 printing blocks on each slide. Random scanning of the printed slides confirmed all the 96 slides were printed successfully.

b. Hybridisation of the microarray slides with wild type and lintless cotton probes

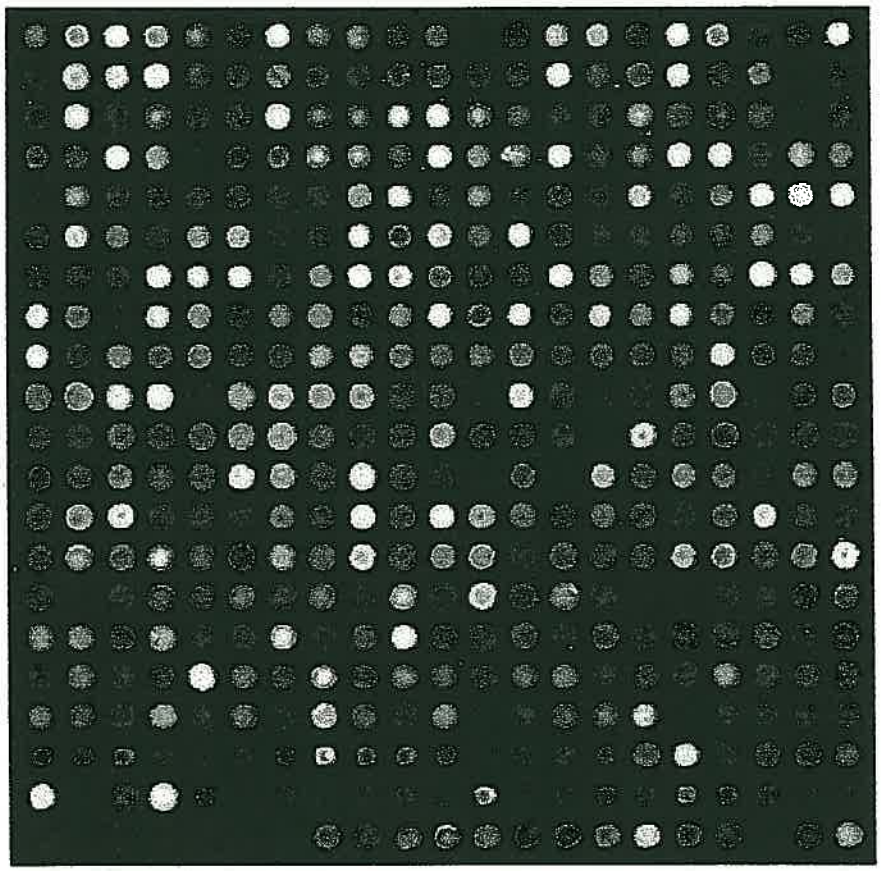
For probing the microarray slides, A lintless cotton mutant 4A-183 was used to compare to a wild type cotton DP-16. Figure 2 illustrates the differences of the 4A-183 2-dpa ovule (right) to DP-16 2 dpa ovule (left). The second day of anthesis, DP-16 ovules are covered with fibre initials, whereas the 4A-183 has only occasional fibre cells initiated.

Fig. 2 Wild type cotton ovule (left) compared to lintless mutant ovule (right) at 2 dpa



Two fluorescent probes were generated by labelling the mRNAs isolated from 0 dpa ovules of DP-16 and 4A-183. The mRNA from DP-16 was labelled with Cy3 fluorescent dye which emits green fluorescence and 4A-183 ladled with Cy5 which emits red fluorescence. The two probes were mixed and hybridised to the microarray slide. After processing, the slide was scanned with a GenePix scanner and image analysed with GenePix software. Figure 3 shows the image of this microarray experiment.

Fig. 3 Microarray image of DP-16 and 4A-183 comparison (one block shown, total 24 blocks on a slide). The yellow dots indicate the gene expression levels of the dots in DP-16 and 4A-183 are relatively the same. The green dots mean the genes are highly expressed in DP-16 but not in 4A-183. The red dots mean the genes are highly expressed in 4A-183 but not in DP-16.



The same experiment was repeated 4 times. In two experiments the probes were labelled as described above. In the other two experiments the probes were labelled the opposite way, eg. DP-16 labelled red and 4A-183 labelled green.

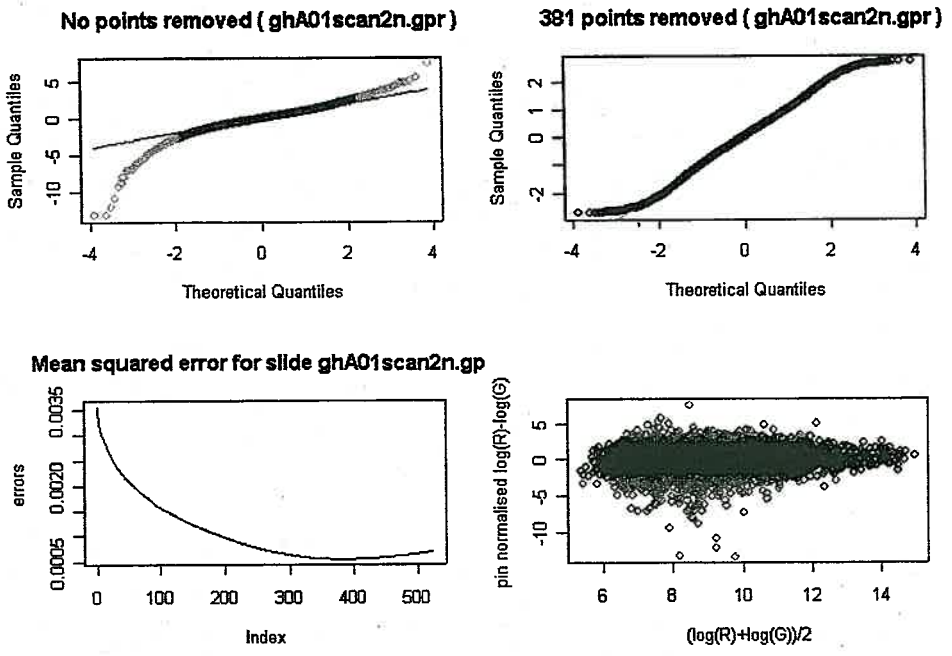
The data generated from the 4 experiments were analysed with a software package tRMA (tools for R Microarray Analysis) developed at CSIRO Mathematical and Information Science. The data from each slide were firstly normalised and differentially expressed genes were identified for each slide (Fig.4). Than the lists of differentially expressed genes from 4 slides were compared. The final output of the analysis was a list of genes that were differentially expressed over all 4 slides. Table 2 lists the number of differentially expressed gene identified with two different normalisation methods compared, namely spatial normalisation and pin normalisation. It also lists the number of differentially expressed genes found in 3 slides out of 4, as well as the number of differentially expressed genes in all 4 slides.

Most of the genes listed as differentially expressed are unsequenced clones, therefore it is necessary to sequence these clones and the sequences will be used to search databases to extract information indicating their function. As to the novel genes (no matches in database), more extensive expression studies (eg. whether their expression is tissue-specific and developmental stage-specific) need to be carried out. To finally confirm the functions of the interesting genes, we need to resort to gene knock out and gene over expression studies using transgenic cotton technology.

Table 2.

Slides	No.DiffExpGenes (Spatial Normalisation)	No.DiffExpGenes (Pin Normalisation)
GhA01	696	381
GhA03	960	801
GhA04	947	736
GhA05	768	507
No.DiffExpGenes Over 3 Slides out of 4	235	155
No.DiffExpGenes Over 4 Slides out of 4	99	67

Fig.4. Find differentially expressed genes using tRMA (red dots represent genes whose expression goes up or down significantly between the lintless mutant and wild type.



6. **Discuss the results, and include an analysis of research outcomes compared with objectives.**

We have produced high quality cotton ovule cDNA microarrays and generated a list of potentially interesting genes in terms of fibre initiation by comparing a lintless cotton mutant to a wild type cotton.

Cotton is a difficult plant to work with in the context of molecular biology. To prepare microarray probes, we need to isolate biologically active RNA from cotton tissues. This seemly routine procedure is made difficult due to interference by high levels of endogenous phenolics, polysaccharides, and secondary metabolites. Extensive efforts were made to try and optimise different RNA extraction methods to generate good microarray probes. The justification of these efforts is evident by the production of the microarray images that are of highest quality.

The lists of interesting genes generated by tRMA are somehow longer than expected. Due to the unknown genetic background of lintless 4A-183, it is difficult to distinguish clearly the differentially expressed genes that are associated with fibre development from those that are associated with genetic background. Sequence information and more detailed expression studies will help to clarify these two groups, but the final confirmation of the association of the genes with fibre development should come from studies of gene knock out and over expression in transgenic cottons.

7. **Provide an assessment of the likely impact of the results and conclusions of the research project for the cotton industry. Where possible include a statement of the costs and potential benefits to the Australian cotton industry and future research needs.**

It is too early to assess the impact for cotton industry at current stage. However, the process of gene discovery will generate new Intellectual Property for Australia. Ownership of this IP will enable us to use it in our variety development program and we will be able to trade access under suitable conditions for return access to other proprietary traits of value for Australian cotton. A better understanding of the molecular processes in fibre initiation and development will allow us to design new strategies to genetically improve fibre yield and quality in Australian varieties.

8. **Describe the project technology (eg. commercially significant developments, patents applied for or granted licenses etc).**

NA

9. **Provide a technical summary of any other information developed as part of the research project. Include discoveries in methodology, equipment design, etc.**

NA

10. Detail a plan for the activities or other steps that may be taken;

(a) to further develop or to exploit the project technology.

This research is further extended by a project titled:" Development of an unigene set of cotton clones for general microarray analysis of gene expression in cotton plants" funded by CRDC (CSP137C).

(b) for the future presentation and dissemination of the project outcomes.

Data generated in the project will be published in peer reviewed Journals when the experiments are completed.

11. List the publications arising from the research project.

The results obtained so far are still too preliminary to publish.

12. Are changes to the Intellectual Property register required?

No

Part 4 – Final Report Plain English Summary

Provide a half to one page Plain English Summary of your research that is not commercial in confidence, and that can be published on the World Wide Web.

As the World's leading natural fibre cotton is an important contributor to Australia's economy. Cotton fibres are single celled hairs on the surface of the seed that have been studied and a microscopic level, but little is known about the genes that regulate the initiation and development of these cells. A better understanding of the molecular basis for initiation and development will allow us to design more rational strategies for cotton fibre improvement using genetic engineering technologies, as well as through conventional plant breeding using molecular marker techniques.

Two different approaches are being used to dissect the control processes in cotton fibre development: a directed approach isolating genes known to be involved in leaf and stem hair development from another plant, Arabidopsis (these hair cells are developmentally similar to the seed hairs on cotton); and a more random, or whole genome, approach by sequencing all the genes expressed during early stages of fibre development and then "mining" these sequences for key control genes using computer analysis, coupled with experimental determinations of when and where each of the many thousands of genes are expressed.

This project is to carry out the expression analysis phase of the cotton fibre genomics project using a newly developed technique of microarray expression analysis. Microarrays are glass slides onto which minute amounts of many thousands of genes are spotted in an ordered sequence using an arraying robot. Many copies of each slide are made and these can then be screened against labelled RNA samples from fibreless mutant and wild type cotton. The expression levels of all the genes on the slides can then be determined simultaneously using a sophisticated scanner and software that can then tell us which genes increase, decrease or stay the same under each of the different conditions. We have amplified 10,410 cDNA clones from our cotton ovule libraries and printed them onto 96 microarray slides. The slides have been used to study the gene expression pattern of a particular lintless cotton mutant as compared to a wild type cotton. A statistical package for microarray data analysis tRMA was used to extract lists of differentially expressed genes and we are in the process of sequencing these genes. In this way we hope to gain valuable information on what genes are important for the initiation of fibres. In the longer term this should allow us to modify important traits such as fibre yield using transgenic plant technologies.