Part 1 - Summary Project Det: Final Report

Report Due Date:	29-Sep-00	C Project Number:	CSP110C

Project Title: (< 15 words) Cotton Fibre EST Sequencing: Gene Discovery Towards Improving Cotton Fibre Quality

Output: Profitibility and Competitiveness

Research Progr: Breeding and Biotechnology

Part 2 - Project Contact Details

Admin Contac: Ms (Title (i.e. Mr/M) | Dianne (First Name) | Rosson (Last Name)

Organisation: CSIRO Plant Industry
(Name of organisation that will be administering the funding)

Postal Addres: 1600 (PO Box) | (Street)

Canberra City (Town) | ACT (State) | 2601 (Post code)

02-624652 (Phone) | 02-624650 (Fax) | D.Rosson@pi.csiro.au (Email)

Primary Resea: Dr (Title (i.e. Mr/M) | Yingru (First Name) | Wu (Last Name)

Organisation: CSIRO Plant Industry
(Name of organisation that will be administering the funding)

Postal Addres: 1600 (PO Box) | (Street)

Canberra City (Town) | ACT (State) | 2601 (Post code)

02-6246501 (Phone) | 02-6246500 (Fax) | Yingru.Wu@pi.csiro.au (Email)

Supervisor: Dr (Title (ie Mr/Ms) | Danny (First Name) | Llewellyn (Last Name)

Organisation: CSIRO Plant Industry
(Name of organisation that will be administering the funding)

Postal Addres: 1600 (PO Box) | (Street)

Canberra City (Town) | ACT (State) | 2601 (Post code)

02-624654 (Phone) | 02-624650 (Fax) | D.Llewellyn@pi.csiro.au (Email)

1. Outline the background to the project.

The cotton fibre is a single cell on the epidermis of the cotton ovule that elongates, fills with cellulose and dries out to form the lint that we harvest for textile manufacture. Little is known about the genes involved in fibre development or the signals that determine which cells on the outer surface of the ovule will become lint or the shorter linters. We have been investigating fibre development in two ways, one by trying to isolate genes from cotton that are like genes in model plants that control trichomes (the hairs on stems , leaves and roots that are anatomically like cotton seed trichomes or fibres) and a second more random or genomics approach of sequencing all the different genes being expressed during early fibre development. By comparing these sequences with other genes in international databases we may be able to identify the types of genes that control patterns of gene expression in fibres or for key enzymes involved in cell expansion and growth. The tens of thousands of genes we will isolate from ovules and fibres can then be used for gene expression profiling using microarrays and this will provide another way of establishing correlations between cotton yield and quality parameters and the expression of particular types of genes that can then be targeted with gene technology to alter the characteristics of the fibres currently produced by the cotton plant.

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

- Prepare cDNA libraries from young cotton ovules at the time fibre cells begin to initiate.
- Sequence 2500 clones from the cotton cDNA libraries.

High quality cDNA libraries were prepared from cotton ovules at two stages –3 to 0 and 0 days which cover the stages when particular epidermal cells of the ovule become committed to becoming fibre cells. The –3 to 0 day ovules were removed from the plant and cultured in vitro in the presence of the protein synthesis inhibitor cycloheximide which we have shown in other systems stabilises rarer messenger RNAs and hence tends to enrich for RNAs oif regulatory genes such as transcription factors. The 0 day library was also normalised – a process which tends to equalise the representation of both rare and abundant transcripts. Approximately 1300 clones were sequenced from the cycloheximide library, 1100 from the 0 day library and about 600 from the normalised 0 day library.

3. Detail the methodology and a justification for the methodology used.

We were interested in finding novel genes that are involved in cotton fibre cell initiation and elongation so that they could be manipnlated in transgenic plants to produce cotton with novel fibre characteristics. One strategy is to use random sequencing of the genes expressed at different stages of fibre development (these are ESTs or expressed sequence tags) and comparing the sequences with the large international databases of genes from all different organisms. Because genes with similar functions often have similar sequence it is possible, for example, to infer functions of unknown cotton genes because of their similarity to an animal, bacterial or plant gene of known function. We were particularly interested in those regulatory genes that might control the important aspects of initiation and growth so we were looked for particular classes of sequences which are known to bind to DNA or in other ways influence the expression of other genes. Ultimately we wanted to use both the sequenced clones and other unsequenced clones from the libraries spotted onto glass slides (microarrays) to explore the expression patterns of large numbers of genes during fibre initiation and development as an alternative way of identifying important regulatory genes. The libraries constructed and seqnences determined will provide us with a valuable resource for all future studies of fibre growth and development iu cotton.

4. Detail results including the statistical analysis of results

The important aspects of the analysis of the cDNA libraries are its quality (size of clones, number of clones with no inserts and the level of redundancy ie. the level of repetition of the same sequences). The three different libraries (CHX – cycloheximide treated ovules, OCF – 0 day ovules, and nOCF – normalised 0 day ovule) were all found to be of high quality in respect of the length of inserts and number of clones containing au insert. All three had relatively long inserts from 1 to up to a few kb in length. The background of clones containing no inserts varied from library to library as indicated in Table 1. The CHX library had the lowest background with fewer than 8% of clones being empty. The OCF library was about twice this and the normalised library was about double again (this is a common feature of normalisation that leads to an increase in the background while removing many of the more abundant clones). The other feature we were interested in was how unique were the clones in the libraries we had constructed. Two other sets of cotton fibre sequences have recently been submitted to the international sequence databases,

on from Prof Ben Burr at the Brookhaven National Laboratory and the other from Prof. Thea Wilkins at the University of California. Both are sequences from about 6-10 days post anthesis so are from fibres during their rapid elongation phase. Some of our sequences overlap with those from these other ESTs but only about 30-40% which means that 60-70% of the sequences in our libraries are novel. Overall about 20-30% of the sequences we have analysed are completely novel and either match proteins from genes of unknown function or do not match any genes currently in the international sequence databases. The challenge now is to determine the possible functions of these genes and to identify which if any have key regulatory roles in cotton fibre development. Using the clones on microarrays will be one way of reducing the number of candidate genes that can be analysed further.
The level of redundancy in the libraries was relatively low with about 15-20% of clones represented more than once making most of the clones different and therefore useful for the construction of microarrays, but also making the high cost of sequencing more economical as the same clones are not sequenced many times.

		CHX Library	OCF Library	nOCF Library
Origin		Cycloheximide treated –3-0 dpa ovules	0 dpa ovules	Normalised OCF clones
CDNA library Quality	# clones sequenced	1272	1149	628
	#with good inserts	1170	988	439
	% good clones	92	86	70
	% background	8	14	30
Redundancy	# Clones represented more than once	261	176	62
	% redundancy	22	17	14
Overlap with other cotton ESTs	% clones matching genes in Burr or Wilkins ESTs	29	46	38
Novel Genes	% clones matching unknown genes in databases	18	17	21
	% clones with no match to other genes	10	7	9
	% novel genes	28	24	30

5. Discuss the results, and include an analysis of research outcomes compared with objectives.
The objective of the project was to generate a resource of genes expressed in early stages of cotton fibre initiation that could be used for genomics approaches to the study of cotton fibre development. We have clearly met these objectives and now have over 2000 sequenced clones and many more thousands of unsequenced clones representing many novel sequences that will provide us with valuable tools to dissect the molecular mechanisms of cotton fibre cell initiation, growth and maturation.

6. 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.
The libraries constructed and the clones sequenced will provide a valuable resource for future studies of the genes involved in cotton fibre growth and development and these will be the resources we can use to manipulate cotton fibre yield and quality parameters. They allow us to move into the next stage of studying

genes that determine how many fibre cells initiate, how rapidly they grow and how they are affected by environmental conditions. We propose to use the newly developing field of genomics (the study of genome wide expression of genes) particularly with microarrays to identify the important genes that control these processes and how they inter-relate. This will in the future allow us to use transgenic plants to alter the expression of key regulatory genes that could modify cotton fibre growth characteristics that determine quality and yield traits.

7. Describe the project technology (e.g. commercially significant developments, patents applied for or granted, licenses, etc).
The database of sequences and the clones themselves are a valuable resource to us as they may contain important genes controlling cotton fibre growth and development and hence may represent novel Intellectual Property for CSIRO.

8. Provide a technical summary of any other information developed as a part of the research project. Include discoveries in methodology, equipment design, etc.
None – standard techniques were used.

9. State the recommendations on the activities or other steps that may be taken to further develop, disseminate, or to exploit the project technology.
The clones generated will be used on microarrays to further our knowledge of key genes involved in cotton fibre development and these studies will be continued in our new project CSP 119C begun in July 2000. We have started to amplify the inserts in the sequenced clones and another 8000 unsequenced clones and these will be spotted onto glass slides using a microarraying robot at CSIRO Plant Industry. These microarrays will be used to examine the changes in expression of all 10000 genes at different developmental stages of cotton fibre development as well as comparisons between expression patterns in normal cotton compared to fibreless mutants of cotton, or between cottons of different quality characteristics.

10. List the publications arising from the research project.
No publications have as yet arisen from this small one year project.

PLAIN ENGLISH SUMMARY

The cotton fibre is a single cell outgrowth from the surface of the cotton seed and is the primary commercial product for which cotton is grown around the world. The yield of fibre and the quality of the fibre (length, strength, uniformity, extensibility, and dyeing characteristics etc.) are all determined by the interaction of the plants genes with environmental parameters. Plant breeders attempt to find the right combination of genes that suit the Anstralian production systems to achieve the highest possible yield and qnality, but with the advent of biotechnology it should be possible to augment this process with more directed genetic changes if we have a better understanding of what genes are involved and how they determine important physical and chemical characteristics of cotton fibres. This project began our use of a genomics approach to try to define some of the important genes involved in cotton fibre production and involved the construction of two different cDNA libraries which represent all the genes expressed in a particular tissues and the sequencing of some of them (over 2500) to gauge the types of genes being expressed. These libraries and sequences will now be a valuable resource for many of our cotton fibre biology projects and will allow us to move to the next phase of gene discovery of using gene microarrays to examine the expression profiles of many thousands of cotton genes dnring fibre development. Such studies will hopefully identify key genes that could be manipulated in transgenic plants to modify cotton fibres for improved or novel characteristics.