REPORT FORMAT

REPORT TYPE: Annual Progress Report: ☐

Final Report: ☒

Part 1 - Project Details

Project Title: (< 15 words) Isolation of gene sequences to manipulate the cotton fibre

Project Number: UA5C

Part 2 - Contact Details

Admin Contact: Ms (Title) Angela (First Name) Thomas (Last Name)

Organisation: Adelaide University
(name of organisation that will be administering the funding)

Postal Address: Research Branch, Adelaide University
PO Box
Adelaide (Town) SA (State) 5005 (Postcode)
(08) 8303 3388 (Phone) (08) 8303 3700 (Fax) angela.thomas@adelaide.edu.au (Email)

Principal Researcher: Dr (Title) Sharon (First Name) Orford (Last Name)

Organisation: Adelaide University

Postal Address: Department of Molecular Biosciences, Adelaide University
PO Box
Adelaide (Town) SA (State) 5005 (Postcode)
(08) 8303 3013 (Phone) (08) 8303 4399 (Fax) sharon.orford@adelaide.edu.au (Email)

Project Supervisor: Assoc. Prof. (Title) Jeremy (First Name) Timmis (Last Name)

Organisation: Adelaide University

Postal Address: Department of Molecular Biosciences, Adelaide University
PO Box
Adelaide (Town) SA (State) 5005 (Postcode)
(08) 8303 4661 (Phone) (08) 8303 4399 (Fax) jeremy.timmis@adelaide.edu.au (Email)

PLAIN ENGLISH SUMMARY OF PROJECT OUTCOMES

Cotton fibres, or lint, are very long single cells which develop during the weeks after flowering. Fibre development is directed by the ordered expression of genes which make signalling molecules, enzymes and structural proteins. Together with the environment in which the plants grow, they are responsible for determining the characteristics of the fibre.

Genetic manipulation of the cotton fibre requires an understanding of the genes which are important in controlling the commercially desirable characteristics of the fibre, a subject which has been addressed in past CRDC-funded research in our laboratory, as well as in project UA5C. Once this information is gained it may be possible to alter fibre characteristics and yield by specifically changing the level and/or timing of the expression of particular genes.

The aims of this project were two-fold. Firstly, we intended to isolate, characterise and test genetic elements which control genes expressed at very high levels in the cotton fibre but at low levels elsewhere in the plant. We obtained seven such elements, three of which were shown to be fibre-specific in cotton transformation experiments. These important elements are now available for further research and engineering of the fibres, as they allow expression of genes (such as pigment or pest resistance genes) just in fibre cells.

A second aim was to target and isolate genes which had not been isolated from cotton, but which, based on data from other plant systems, we suspected would be important in fibre development. We isolated and characterised several such genes, many of which are regulatory, that is, control the expression of other genes, and may have a role in the initiation of cotton fibre growth. This significant preliminary analysis has put us in a position to test these genes, in order to identify the best candidate(s) for manipulation of cotton fibre characteristics and yield.

BACKGROUND TO THE RESEARCH PROJECT

Project UA5C was designed to capitalise on results obtained during the tenure of a CRDC/APRA PhD scholarship (UA1C) and a CRDC Postdoctoral Fellowship (UA4C). Five classes of abundant messenger (m)RNAs had been identified, which are expressed specifically in cotton fibres. All five were shown to be encoded in cotton by gene families of six or more members, thereby complicating the task of isolating the corresponding genes and their regulatory elements. Nevertheless, three genes showing similarity to one of the cDNAs (pFS6) had been isolated, together with one gene corresponding to a second cDNA, pFS18, by screening a Siokra 1-4 genomic library. The four genes had undergone preliminary sequence analysis and the promoters

from three had been subcloned into a GUS expression vector, for testing of their fibre specificity by transient assays.

PROJECT OBJECTIVES

Broadly, the project aims were to:

1. isolate, characterise and test cotton fibre-specific gene promoters and;
2. target and isolate specific genes with documented potential in fibre manipulation.

Primarily, we aimed to follow up our five fibre-specific genes, firstly by testing our existing three promoter::GUS constructs in transient assays and whole plant transformants, and secondly by using a novel PCR method to isolate further promoter sequences, which had proven difficult to identify by conventional means. The new promoter sequences would then also be tested for tissue-specificity in transient assays and whole plant transformants. Results from other plant systems offer possible short-cuts to successful manipulation of the cotton fibre, and the second major aim was to target and isolate characterised genes which we expect will be important in cotton fibre development. Among the named candidates were expansin and XET, both of which allow cell wall expansion. We also aimed to characterise cotton genes which correspond to those involved in *Arabidopsis* trichome development.

These aims were largely achieved to the extent that, in some cases, the results exceeded our expectations. Our progress was rapid despite the fact that we were funded for the salary of one postdoctoral scientist and running expenses. Significant contributions were made by two Honours students and by two PhD students currently in the laboratory (funded by ARC-SPIRT APA(I) and University of Adelaide scholarships). These students were attracted to cotton research by results from CRDC-funded projects. An Honours student is working currently on aspects of cotton fibre gene expression and she and the APA(I) postgraduate are formally cosupervised by both Jeremy Timmis and Sharon Orford. The total additional (non-CRDC) support for this work over the last three years, in the form of stipends and running expenses for students, amounts to approximately $90,000.

METHODOLOGY

The project involved use of a broad range of molecular hiological techniques, as outlined below.

1. Analysis of fibre-specific gene promoters

Further characterisation of the four isolated genes was carried out by sequencing, to show whether the gene corresponded exactly in sequence to the fibre-specific cDNA, or whether it encoded a related protein product. Their promoter regions were to be tested for fibre-specificity, by subcloning the relevant regions into a GUS expression vector, and testing of the constructs by particle bombardment (with a machine to be custom-made within the Department). Positive results would be followed up by transformation of whole plants with the constructs, to be carried out at CSIRO Division of Plant Industry in Canberra. If the promoters were proven to be fibre-specific in whole plant transformants, then a range of different gene constructs would be made and tested by transient assays, using smaller regions of each promoter. Such experiments would define the minimal promoter elements or enhancers required for fibre specificity.

2. Isolation of further fibre-specific gene promoters

It was possible that the four genes previously isolated do not encode the fibre-specific transcripts. In addition, genes corresponding to the remaining three classes of fibre-specific mRNA had not been isolated, since screening genomic libraries with pFS3 and pFS17 yielded no positives, and the expansin had yet to be attempted. A PCR-based approach was to be used to isolate further candidates for fibre-specific genes. The cotton genome is complex, with each gene often present as a family of closely-related members. A PCR-based approach could employ primers specific for each fibre-specific gene and avoid time-consuming library screens.

Once a number of fibre-specific genes had been isolated and their promoters tested as above, it would be possible to compare the promoter sequences in an effort to define the exact regions responsible for the fibre-specific expression.

3. Genes expressed in the developing cotton fibre

A further possibility for the manipulation of fibre characteristics is that genes which are normally expressed in the developing fibre cell may be modified to give novel products, or their temporal expression may be altered. There were four parts to this section of the project:

a. The best candidate from our set of fibre-specific genes encodes expansin, a protein known to be pivotal in the growth of plant cells. The expansin cDNA was to be fully sequenced and corresponding gene(s) isolated and characterised as described above, by genomic library screening and/or a PCR-based approach.

b. A second enzyme known to be involved in plant cell expansion is endo-xyloglucan transferase (XET), and we proposed to isolate cotton XET using a PCR-based method and knowledge gained from known plant XET gene sequences. Published primers would be used to

amplify XET sequences from cotton genomic DNA. The identity of the cloned product could be verified by sequencing, and positive clones used as probes to isolate XET cDNAs and genes from cotton libraries.

c. There are many other examples of genes which could conceivably be of importance in development of cotton fibres and which have been characterised in other plant systems, such as in *Arabidopsis* leaf hair (trichome) development. Such candidate genes would be identified by literature searches and cotton homologues isolated using a PCR-based method, as for XET. Any interesting cotton isolates would be the subjects of further investigation, including established methods such as Southern blot and expression analyses.

d. In addition, it was hoped to carry out a differential screening experiment to identify genes which are turned on once fibre cells enter the secondary wall synthesis stage of their growth. This method was successful in the identification of the original five fibre-specific clones.

DETAILED RESULTS AND DISCUSSION

In addition to the work described below, Sharon Orford attended the ANGIS Bioinformatics course in Adelaide in July, 1997. The results from the work on LTP genes were presented as a poster entitled "Specific expression of two lipid transfer protein genes during elongation of cotton fibres" at the Lorne Genome Conference in January, 1998 and three papers were published during the tenure of this project. In addition to our promoter::GUS constructs, the cotton GL2 cDNA was made available to CRC-funded project 3.3.1 in Canberra, on a collaborative basis, and we also sent two expansin clones to Dr Yong-Ling Ruan (for use in CRDC project CSP87C).

1. Analysis of fibre-specific gene promoters

Further subcloning and sequencing was carried out of the three pFS6 (LTP) genes previously isolated, namely *FSltp1*, *FSltp2* and *FSltp3*, and the LTP-like gene, *FS18A*, following restriction mapping of the genomic clones. A second pFS18 genomic clone was purified and subcloned but sequence analysis of the coding regions revealed presence of an LTP gene, *FSltp2*, rather than a pFS18 homologue. The tissue-specific nature of promoter fragments from *FSltp1*, *FSltp3* and *FS18A* was verified by particle bombardments and transient GUS assays, using an improved protocol. The promoters of the three genes were sequenced in their entirety using subclones and internal primers, to generate over 1 kb of reliable sequence upstream of each of the three genes.

The three constructs were made available for CSIRO Division of Plant Industry for testing in whole plant transformants in April, 1997. A number of transformed cotton lines have been generated for each construct, and results indicate that all three are fibre-specific, with expression of the reporter gene only in the fibres of transformed plants. Seeds have been obtained from both primary transformed lines and self-crossed progeny. There is evidence of both weak and strong expression of the reporter genes in different lines, and these results together show that a major aim of our research has been achieved *viz.* the isolation, characterisation and *in planta* demonstration of cotton fibre-specific gene promoters. The expression level of the reporter gene in each transformed line is currently being quantitated and because the data has not been obtained, we were unable to progress to experiments to identify the "minimal" fibre-specific promoter.

RNA was prepared from all tissues of *G. hirsutum* cv. Siokra 1-4, including freshly-harvested fibres. Samples of the RNA were reverse transcribed and an RT-PCR experiment carried out, using gene-specific primers to determine the expression patterns of the isolated genes. In contrast to the results obtained from the transient and stable transformation experiments, none of the three LTP genes (Orford and Timmis, 2000) or *FS18A* (Orford *et al.*, 1999) exhibited tissue-specific expression patterns, although *FSltp1*, *FSltp3* and *FS18A* transcripts were more abundant in fibre cells than in other cotton tissues. The genomic origin of each gene was determined using the same set of gene-specific primers.

CRDC funds also facilitated the construction of a biolistics apparatus in the department, and the operating parameters of particle bombardment procedure were refined for the routine testing of putative fibre-specific promoters in cotton and tobacco tissues.

The analysis of fibre-specific gene promoters has been complicated by the genome complexity of cotton and the presence of multigene families. Whilst several genes have been isolated and characterised as being potentially fibre-specific, it is often the case that the genes are expressed in other parts of the plant in addition to the fibres, as illustrated above. To target each gene which is specifically expressed in fibre cells, a PCR-based approach was used. For pFS6 and pFS18, the experiment was designed such that the known genes (which appear fibre-specific but do not match the cDNA sequences) were excluded. A commercial (Clontech) "GenomeWalker" kit facilitated the generation of promoter fragments for all five fibre-specific cDNAs, namely pFS3, pFS6 (LTP), pFS14 (expansin), pFS17 (PRP) and pFS18 (LTP-like), and candidate promoters were cloned and sequenced. All were found to closely match the cDNAs, so this approach yielded potential promoters for pFS3 and pFS17 which could not be obtained previously by genomic library screening. In the cases of pFS3, pFS6 and pFS17, a second "walk" was necessary to generate fragments of sufficient length for testing as promoters. The promoters were then amplified directly

from the genome of Siokra 1-4, cloned and fully sequenced. Significantly, the sequence obtained for the pFS6 product was different from the three genes isolated previously (see above) and may correspond to the highly expressed cotton fibre-specific LTP gene.

The sequences of our nine promoters, including those from two expansin genes (see below) were scanned for known (universal) sequence elements and other conserved regions, which may have indicated the presence of control elements. Computer scanning detected universal "TATA" boxes in all except one promoter, but time constraints prevented detailed analysis.

Each promoter, sized about 1 kb, was then directionally cloned into the GUS expression vector. In June 1999, the pFS3, pFS6 and pFS18 constructs were sent to CSIRO Division of Plant Industry for testing in whole plant transformants, followed by the pFS17 construct in November 1999. Due to a malfunction with the plant growth cabinets here in Adelaide, the promoters were, in the end, unable to be tested by particle bombardment. However, we confidently predict that these four promoters will drive fibre-specific gene expression, more strongly than the three previously tested.

2. Genes expressed in the developing cotton fibre

a) The expansin cDNA

The expansin cDNA was sequenced in its entirety and found to be full-length, conceptually encoding a polypeptide of 258 amino acids which showed significant sequence similarity to homologues from other plants (Orford and Timmis, 1998). Probing of a Northern blot with the expansin cDNA resulted in detection of a transcript which was present in fibres but undetectable in other cotton tissues, a result which was confirmed by RT-PCR. The fibre-specific expression pattern, which then yielded us six different fibre-specific cDNAs, prompted further study on the expansin.

The abundance of the transcript was estimated, by probing of a 12 DPA fibre cDNA library, to be the lowest of the six fibre-specific transcripts. Southern analysis revealed presence of a large expansin gene family in the *G. hirsutum* genome and screening of a cotton genomic library with the expansin cDNA identified two distinct genomic clones which were subcloned and the coding regions partially sequenced. However, whilst the two genes putatively encode expansin proteins, they did not match the fibre-specific expansin cDNA in sequence and therefore probably do not encode the fibre-specific expansin transcripts.

In addition, two GenomeWalker PCR products were obtained (as described above) which were very similar in sequence to the cDNA but different from each other and from the two genes obtained previously (by screening of a genomic library). The expansin work, including further characterisation of the two genes and other clones identified in the same screen and assembly and testing of potential promoter/GUS constructs, has been continued by an ARC-SPIRT PhD student, Sarah Harmer, who started work in April 1998 under the joint supervision of Jeremy Timmis and Sharon Orford.

b) MADS box genes in cotton

An experiment not envisaged in the original project application resulted from the chance finding that one or more MADS box genes were expressed in 12 DPA cotton fibres. MADS box genes constitute a family of factors that regulate gene expression in plants. An expanded RT-PCR experiment resulted in the amplification of MADS transcripts from cotton fibres aged 6, 12, 18 and 24 DPA, and the products were cloned as a set of 12 clones per fibre age. Cross-hybridisation analysis divided the clones into four distinct classes and the small representative clones were sequenced, with the finding that the clones indeed encode MADS box homologues. It is possible that MADS box genes could have crucial roles in the developmental aspects of fibre growth and this promising aspect of the project was continued by Honours students in 1998 and 2000, both under the supervision of Jeremy Timmis and Sharon Orford. This year in particular, substantial progress has been made, such that we now have a number of full-length MADS cDNAs from cotton and are in a position to carry out expression analyses.

c) Cotton homologues of *Arabidopsis* genes

Further to the identification of genes involved in the elongation of cotton fibres and following a thorough literature search, an investigation was begun on cotton homologues of genes involved in *Arabidopsis* trichome development. *Arabidopsis thaliana* is used as a model of plant development and the growth of leaf trichomes, which mirrors that of fibre development in the early stages, has been extensively studied at the genetic and molecular level. Degenerate PCR primers were designed from the sequences of two genes known to regulate trichome development, namely *GLABROUS1* (*GL1*), a member of the MYB class of transcription factors, and *GLABRA2* (*GL2*), a member of the homeodomain (HD) class of transcription factors. PCR experiments amplified putative gene homologues from the genome of Siokra 1-4, which were subsequently cloned and sequenced. In addition, the degenerate primers designed from HD sequences amplified a 1 kb band

from the genome of *Arabidopsis thaliana*, used as a control in the experiment, and cloning and sequencing of this product showed it to be *Arabidopsis* GL2.

Four distinct MYB PCR clones were sequenced and all encoded MYB-like proteins which are probably not GL1 homologues. However, two of the clones, which were virtually identical in their coding regions, showed a high degree of similarity to *MIXTA*, a gene thought to be involved in the determination of cell shape. MYB genes in cotton are the subject of a project being undertaken at CSIRO Division of Plant Industry in Canberra and to avoid duplication, the MYB part of this project was not pursued further.

Several cotton HD PCR clones were sequenced and all seemed to encode homeodomain-like proteins. One clone, HD-M91, showed sequence similarity to the GL2 group of HD proteins and was the subject of further analysis. An RT-PCR experiment using the HD degenerate primers detected transcripts in whole ovules and fibres and a large proportion of the ovule product was shown by Southern blot analysis to be HD-M91, suggesting that this gene may be expressed in initiating fibre cells. Cloning and sequencing of the ovule RT-PCR product revealed HD-M91 cDNAs in addition to a second cDNA, GL2-O2, the sequence of which showed even higher sequence similarity to the GL2 group of HD proteins and which is almost certainly the cotton GL2 homologue.

Further analysis of the regulatory genes which had been isolated (cotton *GL2*, HD-M91 and the MADS box genes), centred on the isolation of corresponding full-length cDNAs and on expression analysis. To this end, a cDNA library from whole cotton ovules on the day of anthesis (flowering), at which time fibre cells initiate development, was constructed. In addition, new cDNA libraries for 12 DPA (elongating) fibres and 18 DPA fibres (undergoing secondary cell wall synthesis) were also constructed. Whole ovules from Siokra 1-4 were collected on the day of anthesis, together with fibres aged 12 DPA and 18 DPA. RNA was prepared from all three tissues and poly(A) preps performed. cDNA of high quality was synthesised for all three samples and, despite some unresolved cloning problems, was cloned into an appropriate phage vector. The 12 DPA fibre and whole ovule cDNA libraries were then plated and screened with the regulatory genes described above. The following discussion details experiments on the two clones GL2-O2 and HD-M91.

c)i. The cotton GL2 homologue

Screening of the 12 DPA fibre cDNA library with GL2-O2 identified a 1.7 kb cDNA, GL2(I), which was sequenced to completion. The 1744 bp clone corresponded exactly in sequence to the GL2-O2 clone and contained a poly(A) tail but was found to be partial, missing the 5' end of

the coding sequence. A 5' RACE PCR experiment was carried out using a commercial kit (Clontech) in order to obtain the missing sequence. Four resultant 1.5 kb PCR clones were sequenced and the full-length 2525 bp sequence was compiled after elimination of sequencing and PCR errors. A single long open reading frame was identified, conceptually encoding a 753 amino acid homeodomain protein which showed a high degree (82% similarity and 68% identity) of similarity to *Arabidopsis* GL2. Since the open reading frame starts at nt 2 in the cDNA sequence, a second set of primers was designed and the 5' RACE PCR repeated in an effort to obtain the 5' UTR. A number of products were sequenced but none of them extended the existing sequence.

Database searches using both nucleotide and protein sequences identified a large number of homeodomain proteins from a variety of plants, with highest similarity to GL2 itself and other members of the GL2 subgroup of HD proteins. In addition, our cDNA was virtually identical to two ESTs from 6 DPA cotton fibres (Burr, unpublished), so it appears that we have isolated the full-length cDNA which corresponds to partial sequences in the EST database.

Genomic Southern blot analysis showed that the full-length cDNA hybridised strongly to only two bands in tetraploid cotton, implying that there are few GL2-related genes in the cotton genome. Screening of a Siokra 1-4 genomic DNA library with GL2(I) yielded three clones which, upon purification and restriction, were found to be identical to each other. A 4 kb *Bam*HI fragment was subcloned (as part of project UA6C) and preliminary sequence analysis showed that it contains the gene corresponding to the GL2(I) cDNA, that the gene contains a large number of introns and that the subcloned fragment probably contains at least 1 kb of promoter sequence.

Since Northern blotting proved ineffective (due to the presumably low levels of expression of the corresponding gene) for expression analysis, PCR primers were designed to the GL2(I) cDNA, and an RT-PCR experiment showed that the GL2-like gene is expressed more strongly in whole ovules and elongating fibres than in other cotton tissues. PCR of cotton genomic DNA, using the same set of primers, amplified genes from *G. hirsutum* cv. Siokra 1-2, Siokra 1-4, *G barbadense* cv. Pima S-6 and both the extant diploids *G. herbaceum* and *G. raimondii*. The presence of two amplified bands in tetraploid cotton, which corresponded in size to the single band amplified from each of the diploids, suggested that the primers were amplifying more than one GL2-related gene. The two products were cloned and sequenced to show that one clone closely matched the cDNA, whilst the other was slightly divergent.

At this point it became apparent that similar studies on cotton GL2 homologues were being carried out by Dr Lexi Press, in the laboratory of Dr Danny Llewellyn. Following discussions with CSIRO Division of Plant Industry, in June 1999 we made our cDNA available to the cotton group,

on a collaborative basis, and further experiments on cotton GL2 were not pursued further in this laboratory. It has been proposed that *gl2* and related genes became specialised during plant evolution to regulate aspects related to epidermal cell development, and it may be that the cotton GL2 homologue isolated in this study has a pivotal role in cotton fibre cell development.

c)ii. The HD-M91 clone

HD-M91, whilst being more dissimilar to *Arabidopsis* GL2 than GL2-O2 (section c)i above), remained of interest because of its expression in whole cotton ovules (section c) above). Consequently, analysis of HD-M91 was carried out in essentially the same way as described above for GL2. Probing of the cotton ovule cDNA identified two positive clones, HD(I) and HD(II), which were purified and sequenced to completion with the use of subcloning and internal primers. HD(I), of 2200 bp, corresponded in sequence to HD-M91 but was partial, lacking both a poly(A) tail and the 5' end of an open reading frame. HD(II) proved to be a composite clone, containing 129 bp of an unknown cDNA in addition to 634 bp of a partial cDNA which overlapped exactly with the HD(I) sequence.

Gene-specific primers were designed from the HD(I) sequence, which were then used in RT-PCR and genomic PCR experiments as well as 5'- and 3'-RACE to isolate the ends of the cDNA. The primers amplified a single DNA fragment from all tetraploid and diploid cottons tested, and HD-M91 transcripts were detected in whole ovules and elongating cotton fibres, with lower levels in other cotton tissues and maturing fibres. Four clones, all sized at around 2 kb, were sequenced in the 5-RACE PCR experiment, found to be identical to each other and added 491 bp onto the existing HD(I) sequence. The 3-RACE PCR experiment yielded a clone which only extended the 3' end by 7 bp, but which may contain a poly(A) tail at its terminus. The full-length HD(I) cDNA sequence of 2679 bp contained a single large open reading frame which conceptually encodes a homeodomain protein of 775 amino acids.

The nucleotide and protein sequences showed the highest similarity to ovule-specific and epidermis-specific HD proteins from *Arabidopsis* and orchid, which is in keeping with the observed expression pattern of HD-M91. In addition, our cDNA was virtually identical to two ESTs in Ben Burr's collection of 9000 ESTs from 6 DPA cotton fibres, so once again it appears that we have isolated the full-length cDNA which corresponds to partial EST sequences.

Screening of a Siokra 1-4 genomic library with HD-M91 identified three positives of which one, λGhHD3, hybridised strongly to the probe. A 5.5 kb *Sac*I fragment of λGhHD3 was identified as the smallest, single-hybridising band and was subcloned into a plasmid vector. Sequencing of 2.5 kh of the coding region of the gene (in the normal manner) showed that this is the gene which

corresponds to HD-M91, that our subclone clone includes the promoter region and that the gene contains a number of introns.

c)iii. Cotton TTG homologues

GL1 is thought to act together with a second gene, *transparent testa glabra* (*TTG*), to initiate trichome development in *Arabidopsis*. We attempted to isolate and characterise the cotton homologue of *Arabidopsis TTG* (which was, at that time, unpublished) by collaborative research with Dr Mandy Walker at The Department of Plant Science, University of Cambridge. TTG sequences were amplified from a sample of Siokra 1-4 genomic DNA by Dr Walker and the product cloned in our laboratory. Four clones were sequenced and all showed significant sequence similarity to *Arabidopsis TTG1*. The cotton clones could be divided into two groups, based on sequence, and primers were designed to differentially amplify the two sequences, designated TTG① and TTG⑤. Genomic PCRs and RT-PCR experiments showed that both genes are expressed in whole cotton ovules and elongating fibres, as well as in other tissues, and that the two genes are derived from the same ancestral diploid genome.

Screening of a 12 DPA fibre cDNA library yielded two positive clones, TTG(I) and TTG(II), sized 1.4 kb and 2.2 kb, respectively. Complete sequencing of each clones was facilitated by construction of numerous subclones and design of internal sequencing primers. TTG(II) was found to be a chimeric clone which contained a 500 bp of an unknown cDNA, fused to 1700 bp of a TTG-like cDNA. Both clones contained poly(A) tails and a single large open reading frame. Sequence comparisons showed that the two clones are different from each other, with TTG(I) being identical to PCR clone TTG⑤, but TTG(II) defining a new TTG-like gene from cotton. Table 1 shows the degree of amino acid identity and, in brackets, similarity, between the two cotton TTG-like sequences and *Arabidopsis* TTG1. Database searches using the two cDNA sequences identified a number of WD-40 repeat proteins, in addition to *Arabidopsis* TTG1, and sequence alignments with these characterised proteins allowed the domains and repeats of the cotton TTGs to be defined.

TABLE 1. Amino acid sequence identity and (similarity) between *Arabidopsis* TTG1 and two cotton TTG-like cDNAs.

	Arabidopsis TTG1	Cotton TTG(I)	Cotton TTG(II)
Arabidopsis TTG1	-	79% (89%)	63% (79%)
Cotton TTG(I)		-	62% (79%)
Cotton TTG(II)			-

A genomic Southern blot indicated that there may be more than two TTG-like genes in tetraploid cotton and that they are highly conserved across a range of cotton species but, unlike previous cases, TTGs are not encoded in cotton by a large gene family. Screening of a Siokra 1-4 genomic library with the cotton TTG① PCR clone yielded five genomic clones which, based upon restriction analysis, seemed to segregate into three different types, two of which hybridised strongly to a TTG① probe. PCR analysis of the two strongly hybridising clones suggested that one, λGhTTG(A), encodes the gene corresponding to TTG(I) and TTG⑤, and that λGhTTG(C) encodes a previously unidentified cotton TTG-like gene. The result was confirmed by subcloning and sequencing of the relevant genomic fragments from λGhTTG(A) and λGhTTG(C) (as part of project UA6C). The third genomic clone, λGhTTG(B), possibly contains a further TTG-like cotton gene, but has yet to be characterised.

A GenomeWalker experiment was carried out in order to isolate the genes which correspond to TTG① and TTG(II), which had not been isolated in the cotton genomic library screen. PCR "walks" were performed for each gene in both directions and resultant products cloned and sequenced to confirm their identity to the corresponding isolate. Primers were designed from these sequences, and the two complete TTG① and TTG(II) genes amplified as fragments of 1.9 kb and 1.8 kb, respectively, from the genome of Siokra 1-4. Three independent clones from each amplification were sequenced in their entirety and the sequences compiled, with the elimination of sequencing and PCR errors. A summary of the cotton TTG isolates appears in Table 2, with the degree (%) of amino acid identity and, in brackets, similarity, between the four cotton TTG-like sequences and *Arabidopsis* TTG1 in Table 3.

TABLE 2. Summary of cotton TTG-like isolates

Cotton TTG-like sequence	PCR clone	cDNA clone	Genomic clone
TTG1	TTG⑤	TTG(I)	λGhTTG(A)
TTG2	-	TTG(II)	GenomeWalker PCR
TTG3	TTG①	-	GenomeWalker PCR
TTG4	-	-	λGhTTG(C)
TTG5	-	-	λGhTTG(B)

TABLE 3. Amino acid % identity and (% similarity) between *Arabidopsis* TTG1 and cotton isolates

	Arabidopsis TTG1	Cotton TTG1	Cotton TTG2	Cotton TTG3	Cotton TTG4
Arabidopsis TTG1	-				
Cotton TTG1	79 (89)	-			
Cotton TTG2	63 (79)	62 (79)	-		
Cotton TTG3	80 (89)	87 (93)	63 (78)	-	
Cotton TTG4	62 (79)	60 (78)	95 (97)	61 (77)	-

The cotton TTG-like genes characterised in this study are very similar to each other and to the *Arabidopsis* TTG1 protein, as well as to a number of other WD-40 proteins. Sequence analysis showed that the four sequences formed two pairs, with TTG1 and TTG3 being similar to each other, and TTG2 and TTG4 grouping together. Of these, TTG1 and TTG3 showed the greatest similarity to Arabidopsis TTG1. It is possible that a single gene in *Arabidopsis* corresponds to a number of redundant genes in tetraploid cotton, which may be expressed differentially or have diverse developmental roles which are not necessarily concerned with fibre development. It is impossible, from sequence analysis, to deduce which, if any of these four cotton sequences is the functional TTG1 homologue and to do so requires further study.

A preliminary experiment was carried out by our collaborator, Mandy Walker, in which the two cotton TTG cDNAs, TTG(I) and TTG(II), were tested in a transient assay for TTG activity by complementation of an *Arabidopsis* TTG1 mutant. TTG(I) was tested in four separate experiments, and TTG(II) in three experiments, with six bombardments per experiment, but no positives were observed. The result is not necessarily negative, however. It may be that the cotton TTG constructs only complemented weakly, and it is known that several attempts are required for some constructs to yield positive results.

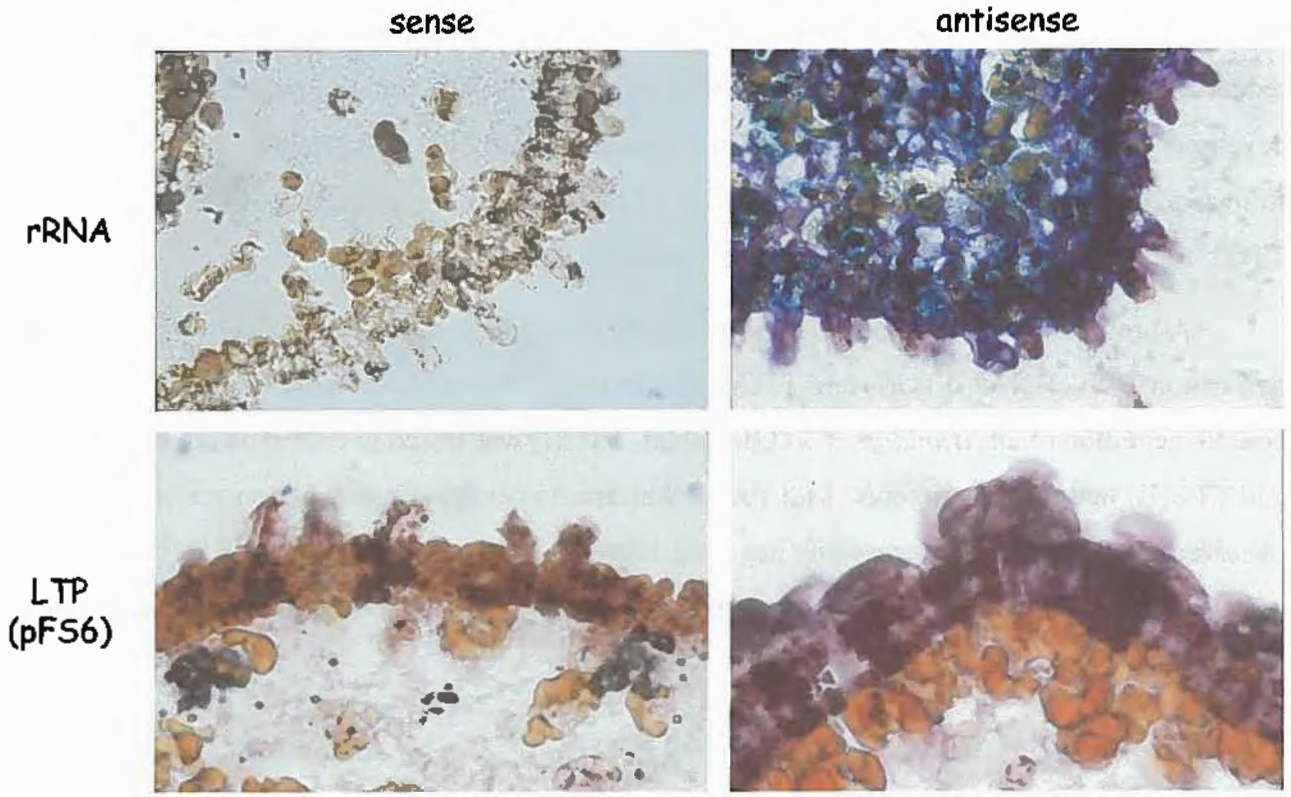
c)iv. RNA *in situs* of cotton ovules

To examine the expression patterns of these promising genes at the cellular level and in early cotton fibre development, an *in situ* method was investigated and tested. Whole cotton ovules aged 0-2 DPA were collected, fixed and embedded in paraffin wax. The tissue was sectioned using resources available in the Anatomy Department and several pilot experiments were carried out, using a method used on *Arabidopsis* ovules and obtained from Dr Jenny Guerin, CSIRO Division of

Plant Industry. Labelled RNA probes were generated for (+) and (-) strands of cotton rRNA, the cotton fibre-specific LTP, HD(I), GL2(I), TTG(I) and TTG(II). In some cases, this involved subcloning and screening of the subclones to generate inserts in the correct orientation for transcription.

Control experiments using the cotton rRNA and the highly-expressed fibre-specific LTP were successful. The rRNA antisense (-) probe hybridised strongly to highly abundant transcripts in all cells, whilst the LTP antisense (-) probe hybridised to a transcript present only in the epidermal cells of the ovule and initiating fibres (Figure 1). RNA *in situ* experiments using the putative regulatory genes were unsuccessful using both this method and a modified technique obtained from Dr Yong-Ling Ruan (CSIRO Division of Plant Industry), but further manipulation of the reaction conditions, in order to accommodate the lower-expressed genes, would be expected to yield results.

FIGURE 1. *In situ* hybridisation of sectioned cotton ovules with rRNA and LTP probes. The negative controls (sense) are on the left, with experimental results on the right.



DISCUSSION OF RESEARCH OUTCOMES

The project objectives for UA5C were largely achieved. In particular, a major aim of our research was completed, such that we now have three proven fibre-specific promoters which, after

further evaluation, will be available for use in aspects of the CRDC Core Biotechnology Program at CSIRO, such as the one to produce indigo-coloured cotton fibres. The promoters will be also be used in further work in this laboratory. In addition, we successfully used a novel technique to isolate further promoters which we confidently predict will be fibre-specific. Together, the set of seven fibre-specific promoters provides a powerful tool in cotton transgenics, enabling any gene to be expressed at predictable levels and in a defined time frame in cotton fibre cells.

In the latter stages of the project, we identified and partially characterised several genes which, on the basis of sequence similarity to known genes, are potentially involved in cotton fibre initiation. Cotton fibre initiation is a poorly understood process at the molecular level yet, if manipulated, could increase the number of initiating cotton fibres per seed and have a direct impact on cotton yield. We are the only group to have initiated a collaboration with Drs Gray and Walker, who cloned the *Arabidopsis* TTG1 gene and are the holders of the patent rights.

The project evolved and expanded somewhat as it progressed, such that some objectives outlined in the initial proposal were carried out by students, who were attracted to the cotton project by results from this and previous CRDC-funded research in this laboratory. Specifically, the characterisation of cotton XET and a differential screen to isolate genes expressed during secondary wall synthesis, were carried out by Nigel Percy as part of research for his Masters degree. The expansin cDNA is probably the most interesting and potentially exciting fibre-specific isolate, but our initial analysis, detailed above, showed it to be encoded in cotton by a large and complex gene family. Further work, to dissect the members of the gene family and investigate the potential of expansin in the genetic engineering of cotton fibres, forms the PhD project of Sarah Harmer. The results from both students have not been covered here.

ASSESSMENT OF INDUSTRY IMPACT AND PROJECT TECHNOLOGY

The primary impact of our work lies in the area of cotton biotechnology and in particular, the genetic manipulation of cotton fibre characteristics. Some exciting results were obtained in project UA5C, with the identification of genes and promoters which have significant potential for altering the morphology or density of cotton fibres. These outcomes should be of interest both to other researchers and to the Australian cotton industry.

In the case of the cotton crop, by far the most important target for genetic improvement is the fibre, which is responding very slowly now to traditional plant breeding methods. Genetic manipulation is the most likely way of successfully producing varieties with, for example, longer, stronger and/or finer fibres. This approach to improvement of crop characteristics is dependent

upon the transferred genes (transgenes) being expressed at appropriate levels in the right tissues and at the correct time. All these features are controlled by specific gene promoters of the type we have isolated for cotton. In the past highly active, constitutive promoters have been used to produce transgenic plants but their use has many disadvantages. For example a transgene expressed at high level in all the tissues of the plant often interferes with normal development and metabolism to effect an overall reduction in plant vigour. This is the reason for commercial interest in our range of genes which are specifically expressed in the fibre.

FUTURE WORK

The full potential of our work should be felt by the industry after only a limited amount of further research and development. In the case of the first three fibre-specific promoters, all that remains is that the whole plant transformants be analysed more fully and the reporter gene expression quantitated. The next batch, of four fibre-specific promoters, must be tested in transient assays, and the results of the whole plant transformants assessed as for the first three experiments. Tissue-specificity must be confirmed in whole plant transformants for all seven promoters before they are considered for use in further research and/or fibre improvement.

Possible commercial impacts of the second part of the work are more distant. Further work on the promising fibre initiation genes would centre on cotton TTG1 and GL2 homologues, but could also be extended to include the MIXTA-like genes and MADS-box genes. The primary aim would be to determine which of these candidate genes has the most potential for the alteration of fibre characteristics, and then to test the effects of changing the gene expression in a test *Arabidopsis* system as well as possibly in whole cotton plant transformants. Initial experiments would include expression analysis by *in situ* hybridisation of cotton ovules, which would be followed up by functional analysis, including complementation of appropriate *Arabidopsis* mutants.

LIST OF PUBLICATIONS

Orford, S.J. and Timmis, J.N. (1998) Specific expression of an expansin gene during elongation of cotton fibres. *Biochim. Biophys. Acta* 1398: 342-346

Orford, S.J., Carney, T.J, Olesnicky, N.S. and Timmis, J.N. (1999) Characterisation of a cotton gene expressed late in fibre cell elongation. *Theor. Appl. Genet.* 98: 757-764

Orford, S.J. and Timmis, J.N. (2000) Expression of a lipid transfer protein gene family during cotton fibre development. *Biochim. Biophys. Acta* 1483: 275-284