Part 1 - Summary Details

CRDC Project Number: **CSP149C**

Project Title: Isolation of novel cotton promoters to drive robust expression of useful genes in transgenic cotton

Project Commencement Date: 1/09/2002 **Project Completion Date:** 30/04/2005

Research Program: On-Farm

Part 2 – Contact Details

Administrator: Graham Brill
Organisation: CSIRO Plant Industry
Postal Address: PO Box 1600, Canberra ACT 2601
Ph: 02-62465431 **Fax:** 02-62465000 **E-mail:** graham.brill@csiro.au

Principal Researcher: Ranamalie AMARASINGHE
Organisation: CSIRO Plant Industry
Postal Address: PO Box 1600, Canberra ACT 2601
Ph: 02-62465294 **Fax:** 02-62465000 **E-mail:** Ranamalie.Amarasinghe@csiro.au

Supervisor: Danny LLEWELLYN
Organisation: CSIRO Plant Industry
Postal Address: PO Box 1600, Canberra ACT 2601
Ph: 02-62465470 **Fax:** 02-62465000 **E-mail:** Danny.Llewellyn@csiro.au

Supervisor 2 Elizabeth DENNIS
Organisation: CSIRO Plant Industry
Postal Address: PO Box 1600, Canberra ACT 2601
Ph: 02-62465061 **Fax:** 62465000 **E-mail:** Elizabeth.Dennis@csiro.au

Signature of Research Provider Representative: ____________________

Part 3.3 – Final Reports (due 3 months after completion of project)

Background

Genetic engineering techniques allow the transfer of novel genetic material from one organism to another and hence have the potential to augment classical plant breeding techniques by extending the gene pool accessible for crop improvement. We already have an extensive program of research aimed at using this new technology to improve the performance of Australian cotton cultivars under our intensive production systems, particularly for enhanced control of insect pests and weeds. This program in collaboration with the Monsanto Company has delivered to the industry new cultivars with enhanced insect tolerance (INGARD and Bollgard II cultivars) and with enhanced tolerance to the herbicide glyphosate (ROUNDUP READY cultivars). The first years of commercial production of INGARD have, however, highlighted some deficiencies in the performance of the insect tolerance of INGARD varieties and this appears to be due to the decline in the expression of the introduced transgene, particularly at the end of the season, but in a significant number of cases much earlier. Similar reports have emerged from the US with both Bollgard cotton and RR cotton. The exact causes of this are still unknown, but other data (Jean Finnegan, personal communication) indicates that steady-state mRNA levels of both the INGARD gene and the selectable marker gene NptII both driven by the 35S promoter of cauliflower mosaic virus or similar viruses are declining in later season plants so it appears that this promoter is being influenced in a negative way by either the environment or the physiology of the plant. These external influences on gene expression are more apparent and economically damaging in Australia because of the high level of tolerance to the insecticidal toxin produced by INGARD cultivars of our target insects, *Helicoverpa armigera* and *H. punctigera*. We cannot change the promoter in the current generation of transgenic plants and the variable performance must be dealt with at a breeding level, but we can develop new promoters that are not as affected by developmental or growth conditions and use these to develop the next generation transgenic cotton plants that will either complement or replace the existing transgenic cultivars. While Monsanto has addressed these early concerns with the development of the two gene Bollgard II and RR Flex cottons there are still some inherent risks in these products that could be improved with better promoters. This project will continued the research begun in CSP102C that had identified a candidate promoter from the small subunit of Rubisco (a photosynthetic gene) as well as use a late season cotton leaf cDNA library to identify other candidates and begin the characterisation of their promoters. The primary objective of these two projects (CSP102C and CSP149C) was therefore to identify and characterise the gene regulatory regions from some cotton genes that might have more consistent patterns of expression throughout the cotton season using both intuitive knowledge about the sorts of genes this might entail or using genomic approaches to carry out large scale expression analyses to identify genes with more robust expression in cotton.

Objectives

To isolate the promoters from cotton genes that are expressed at high levels either throughout the season or, at least in the second half of the season when the viral promoters driving expression of INGARD, and perhaps Roundup Ready, transgenes appear to decline in expression. The project aimed to complete the analysis of a photosynthetic gene promoter identified in CSP102C and to identify other genes that could provide promoters with similarly high levels of expression. These promoters will represent a valuable resource for Australia and could be used to drive useful genes in transgenic cotton irrespective of the source of the transgenic trait so will complement promoters currently available from existing technology providers. They may also have uses in other dicot plants to drive agronomically useful transgenes. The specific objectives were:

Objectives Year 1:

- Complete the production and preliminary analysis of transgenic cotton expressing the GUS reporter gene from the RbcS and cab promoters of cotton.
- Generate T2 seed and select homozygous cotton lines. Analyse the tissue-specific expression patterns of the promoters in glasshouse plants and select the best expressing lines for field testing. Apply for OGTR Licenses for a field trial in October 2003.
- perform further differential screening with LSL library clones in order to obtain more reliable estimates of the types of genes that are more abundantly expressed in late season cotton leaves and/or use microarrays of sequenced cotton ESTs to identify new target genes.
- isolate and sequence the most abundantly expressed gene types indicated by the EST sequence analysis, microarrays and differential screening and analyse their expression in cotton plants by RT-PCR and Northern blotting
- isolate and analyse the promoters of the most interesting target genes using library screening or Genome Walker strategies.

Objectives Year 2:

- Construct promoter-GUS fusion constructs for new candidate promoters and transform into Arabidopsis to verify functionality and begin transformations into cotton. Use transient expression to confirm functionality in cotton.
- Analyse expression patterns of promoter GUS fusions in transgenic Arabidopsis T1 and T2 plants.
- Collect field samples from RbcS GUS and possibly cab GUS transformants and analyse expression levels relative to 35SGUS constructs grown under the same conditions.
- Analyse expression patterns in T0 plants of new promoter GUS constructs and identify lines for subsequent field-testing.

These objectives were achieved with the isolation and characterisation of a small subunit promoter in both Arabidopsis and cotton; the sequencing and bioinformatics analysis of two thousand cotton late season leaf ESTs; the printing of a late season leaf microarray and hybridisation with labelled cotton leaf mRNA from early and late in the season to identify robustly expressed genes. These genomic analyses identified both housekeeping genes, photosynthetic genes and some novel genes that might be a useful source of promoters. The Small subunit promoter was linked to a GUS reporter gene transformed into cotton and homozygous T2 plants have undergone testing both in the glasshouse and under field conditions but we are still analysing the samples and data generated in the field trial that has only recently been completed.

Methodology

This project used standard molecular biology techniques to isolate and characterise the expression of genes identified in our genomic studies. This included isolation of total and mRNA from both glasshouse and field grown cotton, cloning of genes and promoters from both cDNA and genomic libraries of cotton; construction of expression cassettes and transformation into both cotton and Arabidopsis using Agrobacterium mediated transformation protocols. Transgenic cotton plants were generated and expression of the Small subunit promoter analysed by following the expression of a linked reporter gene (GUS or beta glucuronidase) both enzymatically, histochemically and by Northern gel analysis. These will not be described here. The genomic techniques are novel so are detailed below.

cDNA libraries are a collection of DNA clones in plasmids of all the expressed genes present in a sample of mRNA isolated, for example, from a leaf. A highly expressed gene would constitute a larger proportion of the mRNA in the sample so will be cloned multiple times. The relative abundance of any particular sequence within the library of clones is therefore an indication of the abundance of that gene. As we constructed a cDNA library from late season cotton leaves the sequencing of randomly selected clones and a bioinformatics analysis of the relative abundance of individual DNA sequences can be used to tell us the expression level of that gene in late season leaves and so allow us to target particular types of highly expressed genes. A sequenced cDNA clone is called an Expressed Sequence Tag (or EST) and represents a section of the sequence a few hundred base pairs long that can serve as a tag for the gene and be manipulated by a \variety of computer programs for sequence analysis and comparison. In this project we generated about 2000 ESTs from late season cotton leaves and analyse the abundance of the most highly expressed types of genes.

Microarrays are another way of examining expression patterns of genes in a global way. Microarrays are formed by robotically depositing specific fragments of DNA at indexed locations onto microscope slides. The DNA fragments can originate from a variety of sources including anonymous cDNA clones, EST clones (ie sequenced cDNA clones), anonymous genomic clones, synthesized oligonucleotides, or DNA amplified from open reading frames (ORFs) found in sequenced genomes.

Once produced, the microarrays are hybridised with fluorescently-labelled mRNA-derived probes and the bound probes on the array are then excited by light. The fluorescent signal emitted from each spot is a reflection of the abundance of the corresponding sequence in the original probe and hence tissue from which the original RNA was extracted. Microarray technology is ideally suited for making pair-wise comparisons of samples. Two fluorescent tags – often just refereed to as red and green 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 (red and green channels) reflects the ratio of the abundance of that mRNA species in the two probes.

Microarrays are one of the most powerful tools that have recently been developed to bridge the gap between sequence information and functional genomics. The power lies in its scale, sensitivity and the quantitative nature of the data generated. 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 on cotton genomics was only made possible by the installation at CSIRO a few years ago of a robotic microarrayer and slide scanners. As part of this project we constructed a small leaf microarray from about 1800 of the cotton leaf ESTs that were from late season leaves and hybridised it with labelled RNA from early and late season leaves to identify which genes were more highly expressed late in the season or at least that were relatively highly expressed both early and late. This confirmed some of the EST sequence analysis data and suggested new targets from which we might isolate promoters for cotton biotechnology.

Results

Construction of a Late Season Leaf cDNA library and preliminary sequence analysis.

A high quality directional cDNA library was prepared from leaf RNA extracted from field grown transgenic Bollgard II/Roundup Ready cotton plants at a time when the efficacy of their 35S promoter driven Cry1Ac gene against *Helicoverpa armigera* would normally have declined significantly (not shown). 2000 randomly selected clones were sequenced from the 3' end. Approximately 1810 high quality sequences were recovered and clustered into 1340 contigs each representing a different gene type. BLAST searching was used to identify the most closely related entry in SwissProt protein database or where there was no significant hit, to the Genbank Non-Redundant nucleotide database in order to classify the genes by a possible function. These sequences will be made available publicly after the completion of the project and the different contigs are available from the cotton genomics database at the Arizona Genomics Institute using the PAVE sequence browser (http://www.agcol.arizona.edu/pave/cotton/) as the Gh_LSL cDNA library clones (with restricted access at this stage).

Table 1. Abundance of ESTs in the Late Season Leaf (LSL) cDNA library six different clusters with 3 or more ESTs but no significant hit with MEGABlast or BlastN not included.

ESTs	Clusters	Match
253	1	26S ribosomal RNA gene
69	15	RbcS (G. hirsutum)
23	9	Cab Precursor
11	9	EF-1- alpha
11	8	glyceraldehyde 3-phosphate dehydrogenase
9	7	alpha tubulin
8	7	beta-tubulin
7	3	Glycine Rich protein
7	3	Thiazole biosynthesis protein - chloroplast
7	3	PLAZ-POPNI
7	2	germin-like protein
6	4	Fructose bisphosphate aldolase
4	1	PDA-RICCO
4	1	BRU1 - brassinosteroid regulated protein
4	2	Malate dehydrogenase
4	2	actin
4	2	PS II protein
3	1	caffeoyl methyltransferase
3	1	NptII
3	1	18S Ribosomal RNA gene
3	2	ascorbate peroxidase
2	1	EF-2

In order to define the most highly expressed genes at this late stage of growth, the contigs were assembled into higher order groupings based on the identity of the BLAST hits of their consensus sequence. ESTs were counted together if the contig they belonged to had the same best match. It was assumed that the number of ESTs in a group correlated with the expression level of the corresponding gene(s). No distinction was made at this stage between different members of a multigene family. The top twenty assemblies and the corresponding number of total ESTs are shown in Table 1. The most abundantly expressed gene in late season leaves appears to be the 26S ribosomal RNA gene, but as expected, the two photosynthetic genes ribulose bisphosphate carboxylase small subunit (RbcS) and chlorophyll a/b binding protein (Cab) were both highly represented amongst the ESTs from late season leaves, ranking second and third most abundant. House keeping genes such as elongation factor 1, and tubulin were also highly represented. Actin was present, but less abundant than these other housekeeping genes. A number of structural protein genes were represented in the top assemblies as well as some genes for enzymes of primary and secondary metabolism. Many of the more abundant classes were comprised of a large number of contigs, indicating that they were probably not from single genes, but from related, but different members of multigene families. Interestingly, the neomycin phosphotrasferase (Npt II) gene (used as a selectable marker in the commercially released transgenic cotton from which the library was made) also appears in the top twenty most abundant assemblies. The commercial transgenic plants contained four copies of a 35S promoter driven NptII gene (two from the Cry1Ac insert of Bollgard II and two from the Roundup Ready insert).

To get a more realistic picture of the temporal differences in expression of the late season ESTs they were printed to glass slides and probed with labelled cDNAs from different times during development of cotton from early vegetative to during fruit development and maturation.

Analysis of the expression pattern of cotton leaf ESTs during the life cycle of cotton using cDNA microarrays.

Since we were primarily interested in assessing the robustness of gene expression profiles at different developmental ages under field conditions, leaf samples were collected from cotton plants grown in the field rather than in a glasshouse. RNA isolated from young pre-flowering, vegetative plants (6 weeks) was compared on the leaf microarray to RNA isolated from relatively mature plants (20 weeks and 22 weeks) when the plants were carrying their peak boll load, or starting to open their bolls, respectively.

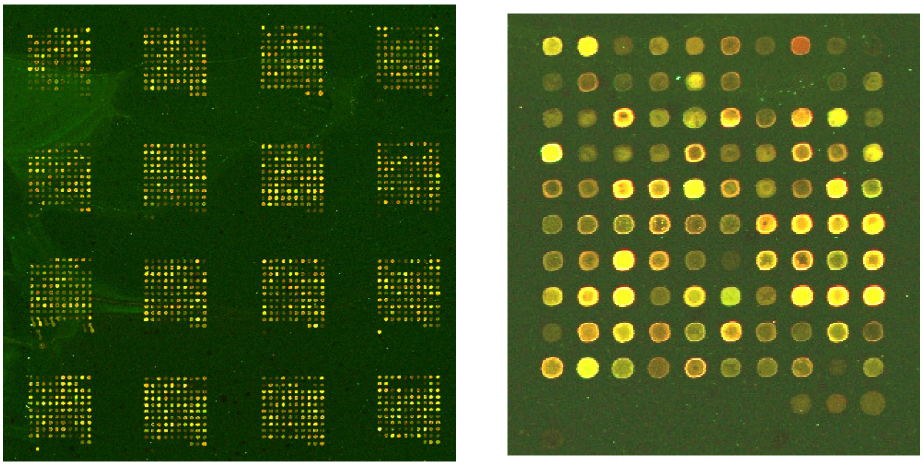

Figure 1. Microarray hybridisation of fluorescently labelled cDNA from leaves of both young and mature cotton plants to a boutique array of LSL EST clones (inset shows one panel at greater magnification). Redish dots represent genes that are up-regulated in the later stage. Greenish dots represent genes that are expressed more in the early season leaves. Yellow dots are expressed equally at both stages of growth.

Table 4. ESTs with the highest average fluorescence over 6-22 weeks

EST	log2R+log2G/2	Identity (tBlastX)
LSL008E11	14.5876	no ID (poor sequence??)
LSL011E08	14.4931	Ribulose bisphosphate carboxylase small chain
LSL008D07	14.4225	Ribulose bisphosphate carboxylase small chain
LSL010F03	14.3445	Ribulose bisphosphate carboxylase small chain
LSL011C09	14.2188	no ID (poor sequence??)
LSL010F04	14.1782	Desacetoxyvindoline 4-hydroxylase??
LSL008B11	14.1542	Ribulose bisphosphate carboxylase small chain
LSL006G03	14.1522	Ribulose bisphosphate carboxylase small chain
LSL011B10	14.1106	Arabidopsis hypothetical protein
LSL008B03	14.0943	no match
LSL002G06	14.0444	Ribulose bisphosphate carboxylase small chain
LSL008G03	14.0248	no ID (poor sequence??)
LSL008E12	13.9191	no ID (poor sequence??)
LSL011A01	13.9863	no ID (poor sequence??)
LSL024C10	13.9801	no ID (poor sequence??)
LSL024H01	13.9560	18 ribosomal RNA
LSL011H06	13.9543	Ribulose bisphosphate carboxylase small chain
LSL006A01	13.9456	no ID (poor sequence??)
LSL005G10	13.9324	Ribulose bisphosphate carboxylase small chain
LSL025B11	13.8915	26S ribosomal RNA
LSL002E06	13.8877	Ribulose bisphosphate carboxylase small chain
LSL007B09	13.8381	Ribulose bisphosphate carboxylase small chain
LSL011A09	13.8249	no ID (poor sequence??)
LSL022E07	13.8246	no match (short sequence)
LSL005C03	13.7900	Ribulose bisphosphate carboxylase small chain
LSL008A03	13.7874	Ribulose bisphosphate carboxylase small chain
LSL025G03	13.7819	Ribulose bisphosphate carboxylase small chain
LSL009H10	13.7113	Ribulose bisphosphate carboxylase small chain
LSL026A08	13.6923	no ID (poor sequence??)
LSL024E02	13.6801	no match
?? Indicate low homology matches		

Each comparison was replicated 4 times as two biological replicates (RNAs from the same plant variety from different fields sampled at the same times) and two dye swaps serving as technical replicates. An example of the hybridisation to the glass slide is shown in Figure 1. tRMA (a statistically based microarray analysis package developed by CSIRO) was the microarray analysis tool used to normalise the raw data (so that different array experiments could be compared) and to identify the statistically significant differentially expressed genes at these time points.

Surprisingly, there were only a small proportion of the genes that were identified to be differentially expressed at the later stages of growth when compared to younger vegetative plants (44/1500 at 20 weeks and only 11/1500 at 22 weeks), the majority being down regulated at the later stages. Tables 2 and 3 (at end of document) show the genes that were up- or down-regulated in cotton leaves aged 20 and 22 weeks, respectively at the 0.001 confidence limit. At 20 weeks only two genes were up-regulated compared to their expression level at 6 weeks. These two ESTs were found to be homologous to a chlorophyll

A/B binding protein and an Arabidopsis thaliana BAC clone of unknown function. At 22 weeks all 11 genes were down-regulated (Table2). The EST, LSL001DO9, homologous to a possible tannin biosynthetic gene, showing the highest drop in expression, 10 times compared to its level at 6 weeks of age.

Since there were few genes that increased expression over time we focussed on those genes that were consistently strong throughout the life cycle of cotton. The log2 transformed average fluorescence values in both the red and green channels were used as an approximation for the overall level of expression of the genes on the microarray at the different growth stages. 6 week old plants were compared against 10 wk, 12 wk, 20 wks and 22 wk old plants from the field on separate arrays with 2 (10 &12 w) or 4 (20 & 22 w) replicates.

Table 4 summarises the ESTs with highest log 2 transformed average fluorescence value (13-14.9) for the whole life cycle (6-22 weeks of age). The twelve ESTs that had the strongest expression levels (>14) at all the times tested included the Rubisco small subunit genes and a few other novel genes. 300 ESTs had expression levels higher than 11 (not shown). The known house-keeping genes ubiquitin and beta tubulin had fluorescence values between 11 and 12 (not shown). These data also confirm that the small subunit gene is a good candidate for isolating a very robust promoter.

Verification of the microarray data by Northern blotting.

To examine the expression levels of the candidate genes in cotton at various stages of growth, and to confirm our microarray data, Northern analysis was performed on total RNA isolated from field grown cotton plants at the similar stages of growth as used for the microarray analyses.

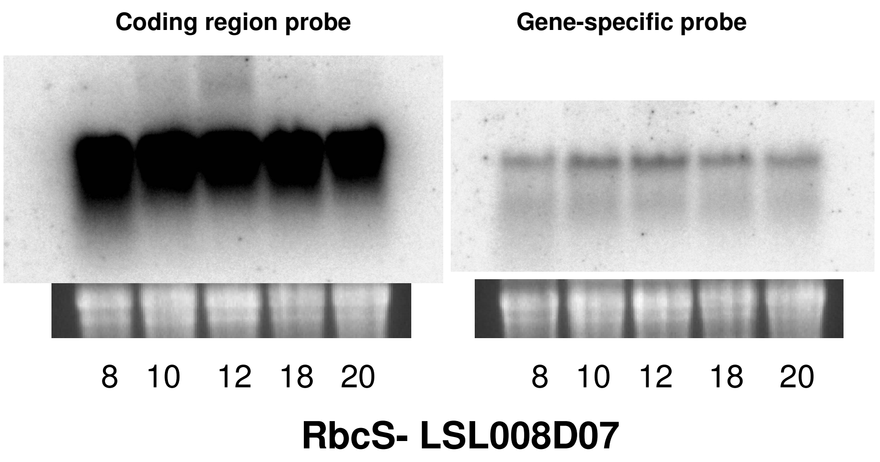

Figure 2. Northern blot of RbcS mRNA levels over time in field grown cotton using either a general coding region probe that would hybridise to all small subunit genes or a gene-specific probe that would just detect LSL008D07

Both generic coding region probes and sequence specific riboprobes made against the EST LSL008D07 (RbcS) revealed that this particular RbcS gene's expression levels did not change markedly over the growth season (Figure 2) and was expressed significantly higher than that of the house keeping gene, Ubiquitin (LSL028G08) throughout the life cycle (not shown). Similar experiments using sequence specific riboprobes made against the ESTs LSL030E06 (Cab) and LSL001D09 (leucoanthocyanidin reductase) revealed differential expression patterns for these two genes as predicted by the array data. The Chlorophyll a/b binding protein is up-regulated at 20w and 22w compared to week 6 confirming the array data, while the leucoanthocyanidin reductase gene is down regulated with time again being consistent with our array results. Of the half dozen genes we verified in this way all varied in the same manner over time as suggested by the microarrays and we feel confident that the array data was provided reliable estimates of expression changes (with the proviso that some cross-hybridisation between different closely related members of gene families may not be distinguished on the arrays).

Taking all the data together, the small subunit of Rubisco would seem to be an ideal target and this was investigated further. Other genes worth investigating were Cab, elongation factor 1-alpha and glyceraldehyde phosphate dehydrogenase but due to time limitations these were not progressed as far as the RbcS promoter.

Isolation of the promoter region of the Rubisco single subunit gene of cotton.

Rubisco small subunit gene was one of the strongest on the array and its cDNA appeared very often amongst the clones represented in a late season leaf cDNA library so was a good candidate from which to isolate a strong promoter. The EST LSL008D07 which was among the highest in expression through out the season was identical in its sequence to a cotton RbcS small subunit gene sequence (Accession X54091) already available in the Genbank database although its promoter sequence was only 520bp long and probably too short for a good promoter fragment for expressing gene constructs. We therefore used the full length LSL008D07 EST to screen a cotton genomic library to isolate the corresponding Rbcs gene from cotton along with a 1829bp long 5' promoter region that we believed would contain all the sequence motifs necessary for proper expression in transgenic plants. The cotton promoter sequence was compared against some other known RbcS sequences from other plants (coffee, potato, Chrysanthemum, maize, and rice) by the ClustalW sequence alignment and comparison program. Closest homology (other than the existing cotton sequence) was shared with the Chrysanthemum gene and six conserved motifs identified by the program MEME were found in all of these sequences, suggesting that the cotton promoter fragment should be functional, although we were keen to verify this in a transgenic system other than cotton because of the long timeframes fort generating transgenic cotton plants.

Testing the functionality of the Rubisco (RbcS) promoter in Arabidopsis

The long 1829 bp promoter fragment and several 5' deletion constructs of the promoter were fused to the GUS reporter gene in plant expression vector (Figure 3) and used to transform the model plant *Arabidopsis thaliana* using the floral dip method. Transformants of the RbcS-GUS vector were selected on media containing Kanamycin and homozygous T2 transgenic Arabidopsis lines produced and analysed for the expression of the GUS gene both histochemically and in quantitative assays. An Arabidopsis line where the GUS reporter gene was driven by the 35S promoter of the cauliflower mosaic virus was used as the control in this experiment. Histochemical staining for GUS gene product showed the activity of the GUS gene under the Rbcs promoter in transgenic Arabidopsis plants.

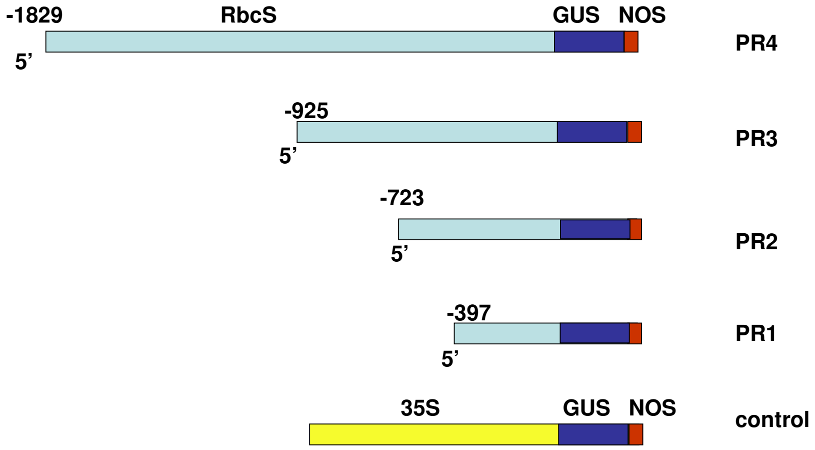

Figure 3.
Promoter GUS fusion constructs introduced into Arabidopsis to test the functionality of the RbcS promoters

All four Rbcs-GUS constructs were functional in Arabidopsis. Figure 4 shows the expression of GUS gene in Arabidopsis plants under the RbcS promoter compared to the 35S promoter of the cauliflower mosaic virus. Unlike the 35S promoter which expresses in all the plant parts, Rbcs promoter drives expression only in green tissues, and is thus is tissue specific.

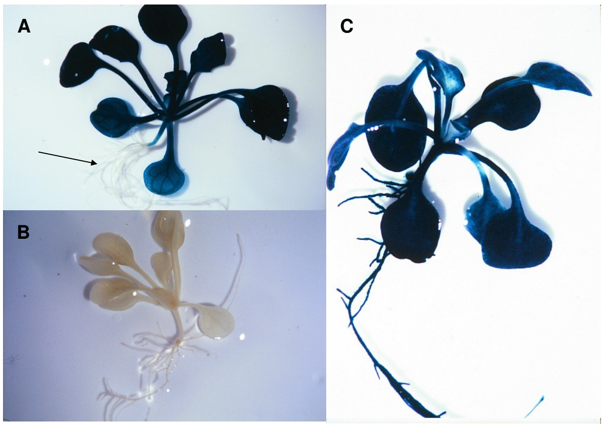

Figure 4. GUS staining patterns in PR4 (RbcS-GUS) (**A**); Non-transformed (**B**) And 35S-GUS (**C**) Arabidopsis plants. Note absence of GUS in roots of PR4 plants (arrow)

To quantify and compare the level of GUS activity in different Arabidopsis transgenic lines, a GUS assay was performed using the fluorometric substrate MUG (Methylumbelliferyl-glucoronide). Figure 5 shows the mean GUS enzyme activity level in Arabidopsis T2 lines containing the promoter constructs PR4, PR3, PR2 and PR1, varying in size for the promoter from the longest at 1829bp to the shortest at 397bp. The highest GUS activity was seen in the PR4 line where the GUS gene was under the longest Rbcs promoter fragment. GUS activity driven by the 35S promoter was only 50% of the activity seen in PR4 plants, suggesting that this long RbcS promoter should be a good promoter for driving transgene expression in dicot plants like cotton. Progressive deletion of the promoter reduced overall activity. Thus, although we did not study the importance of different functional motifs in Rbcs promoter in detail, deletion of at least some of them, if not all, seem to have an impact on the activity of the promoter confirming that it would be best to introduce the longest promoter into cotton.

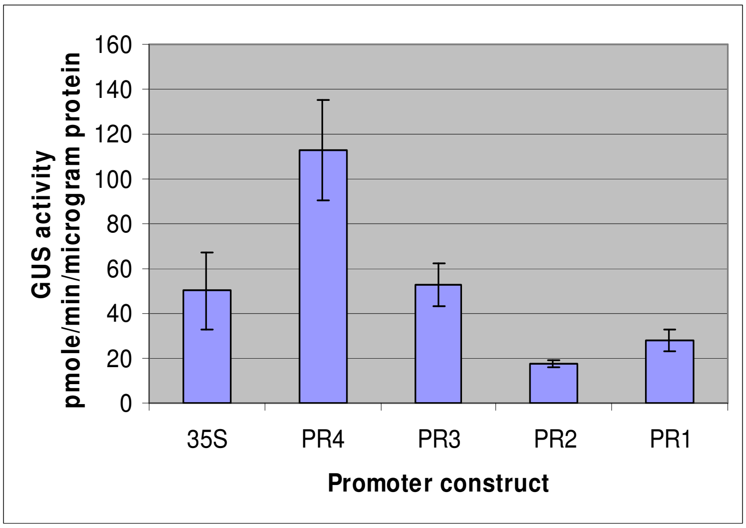

Figure 5. GUS activities of T2 homozygous *Arabidopsis* lines transformed with the full-length RbcS promoter construct (PR4) and various deletions (PR1-3) compared to a 35S-GUS construct. Values are the average of the highest six independent lines for each construct and error bars are the standard deviation.

Generation and analysis of transgenic cotton plants containing Rbcs-GUS construct

To assess the performance of Rbcs promoter in cotton, variety Coker 315 was transformed with the PR4 Rbcs-GUS construct with GUS gene under the 1829bp long Rubisco promoter. Twenty fertile primary transformants representing fifteen different transformation events were generated. Southern analysis of T0 plants revealed successful integration of the GUS gene into the cotton genome and the number of transgene insertions to vary from 1-8. 4 lines had single insertions, 2 lines had two insertions and 14 lines had multiple insertions of the RbcS-GUS gene.

24 T1 seeds from each T0 line were grown to produce T2 seed. Individual T1 plants were screened by GUS staining of leaf tissue to allow the nulls to be removed. Segregation ratios predicted single locus insertions in six lines while the rest of the lines were appeared to have multiple locus insertions. Southern blotting on DNA isolated from 12 best GUS expressing plants from each line confirmed the stable inheritance of GUS gene in T1 generation.

24 T2 seeds from best GUS expressing T1 lines were planted and the zygosity of their parents assessed. Selected homozygous lines were further used in expression studies and for a field trial carried out in 2004-2005 at Narrabri.

Expression of Rbcs promoter in cotton

To determine the specificity of expression of the Rbcs promoter in cotton, we stained different plant parts harvested from the T1 lines. As expected, leaves had very high expression of GUS, while roots had almost no detectable expression. Young stem and the young floral buds also had low levels of expression. These results are consistent with the previous studies and show that the Rbcs promoter activity is restricted to photosynthetic tissues. This was confirmed by Northern blotting using RNA extracted from the different tissues (Figure 6.). Preliminary evaluation suggested that RNA levels did not decline drastically over the growth cycle of the cotton in the glasshouse but more detailed assessment was to be carried out in the field.

Table 2. Differentially expressed genes in leaves of 20 versus 6 week old plants

Name	Average log2 red/ log2 green channel	StdDev	Backtransformed Ratio	Identity
LSL030E06	0.8728184	0.2287398	1.8312369	Chlorophyll a/b protein
LSL007E01	0.7115634	0.0542721	1.6375777	no ID (poor sequence??)
LSL010D04	-0.6996253	0.1829991	0.6157321	Rubisco Activase
LSL026B02	-0.7450087	0.3151972	0.5966643	Anther-specific proline rich protein
LSL021E10	-0.760856	0.1371209	0.5901461	Glycine decarboxylase
LSL024B11	-0.7868149	0.1464028	0.5796223	monodehydroascorbate reuctase
LSL026F02	-0.8102749	0.2092345	0.5702732	vacuolar pyrophosphatase
LSL021D01	-0.8116222	0.2250769	0.5697409	ATP-citrate synthase
LSL023A02	-0.8156689	0.0939586	0.568145	Chalcone synthase 1
LSL028F07	-0.8288328	0.0951687	0.5629846	S-adenosylmethionine synthetase 1
LSL009A04	-0.8370474	0.1755413	0.559788	phospholipid hydroperoxide glutathione peroxidase
LSL023F05	-0.8410112	0.2260474	0.5582521	Acetyl-CoA carboxylase 2
LSL010G12	-0.8510676	0.1420346	0.5543743	Heat shock protein
LSL009G12	-0.8668209	0.0867674	0.5483539	peroxisomal-coenzyme A synthetase
LSL007F03	-0.8700485	0.1583138	0.5471285	Glyceraldehyde-3-phosphate dehydrogenase A
LSL001H10	-0.8815989	0.3126245	0.5427656	S-adenosylmethionine synthetase 1
LSL021F11	-0.8832475	0.4733352	0.5421457	Flavonoid 3',5'-hydroxylase 2
LSL001G08	-0.8902163	0.1235683	0.5395332	Putative peroxisomal-coenzyme A synthetase??
LSL024A05	-0.9048312	0.1006074	0.5340952	phospholipid hydroperoxide glutathione peroxidase
LSL009H11	-0.9159198	0.2447228	0.5300059	Chalcone--flavonone isomerase A
LSL006E01	-0.9234897	0.1202244	0.5272322	Glyceraldehyde-3-phosphate dehydrogenase A
LSL022E10	-0.9484475	0.1092305	0.5181898	Ubiquitin
LSL001D01	-0.9724968	0.1220008	0.5096233	RuBisCO activase
LSL007A07	-0.9777872	0.0713486	0.507758	no ID (poor sequence??)
LSL006B08	-0.9799359	0.1676047	0.5070023	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
LSL030D12	-0.9828274	0.1413171	0.5059871	no ID (poor sequence??)
LSL010H03	-1.0153336	0.0485957	0.4947139	Heat shock cognate 70 kDa protein 1
LSL001A12	-1.0165022	0.1358308	0.4943134	Putative cell wall protein precursor??
LSL003D06	-1.025762	0.1722609	0.4911508	Chalcone synthase 1
LSL022D11	-1.0259435	0.221269	0.491089	RuBisCO activase
LSL030G08	-1.0379777	0.1341095	0.4870097	no ID (poor sequence??)
LSL024A10	-1.0467824	0.2502495	0.4840465	Brassinosteroid LRR receptor kinase precursor ??
LSL022C09	-1.0701763	0.1296061	0.4762608	Early light-induced protein, chloroplast precursor (ELIP)??
LSL007C04	-1.0816504	0.1803082	0.472488	Probable pyridoxin biosynthesis protein ER1 (ethylene-inducible)
LSL031H07	-1.0910944	0.2734936	0.4694052	Probable pyridoxin biosynthesis protein ER1 (ethylene-inducible)
LSL008F10	-1.1200288	0.2134127	0.4600846	Alanine aminotransferase 2 ??
LSL005G02	-1.2330285	0.1584737	0.4254235	Phosphoprotein??
LSL007C10	-1.2468921	0.1897778	0.4213549	Zinc finger protein??
LSL023F08	-1.2663942	0.2956872	0.4156974	no match (Novel gene)
LSL007E05	-1.2692183	0.1144691	0.4148845	cell wall protein precursor ??
LSL025E09	-1.3282378	0.0098551	0.3982544	Heat shock protein ??
LSL004A12	-1.6797477	0.184268	0.3121372	Catalase isozyme 2
LSL001D09	-2.4300366	0.1168966	0.1855607	Leucoanthocyanidin reductase

Up regulated genes are in yellow; ?? Homology low so ID uncertain

Table 3. Differentially expressed genes in leaves of 22 versus 6 week old plants

Clone Name	log ratios (Red/Green)	Std Dev	Backtransformed Ratio	Identity
LSL024A10	-1.2875	0.31603	0.4097	Brassinosteroid LRR receptor kinase??
LSL026B02	-1.1951	0.10644	0.4368	Anther-specific proline rich protein
LSL028C09	-1.3110	0.32605	0.4030	Glutathione S-transferase
LSL030D12	-1.5300	0.67417	0.3463	no ID (poor sequence??)
LSL003D06	-1.2390	0.18361	0.4237	Chalcone synthase
LSL004H10	-1.2190	0.33004	0.4296	NptII (kanamycin resistance gene)
LSL001D09	-2.9551	0.20676	0.1290	Leucoanthocyanidin reductase
LSL005G02	-1.3180	0.19133	0.4011	Phosphoprotein??
LSL009A08	-1.1088	0.60310	0.4637	NptII (kanamycin resistance gene)
LSL009H11	-1.4161	0.14240	0.3747	Chalcone--flavonone isomerase
LSL022C09	-1.5664	0.26667	0.3376	Early light-induced protein, chloroplast precursor (ELIP)??
?? Indicate low homology matches				

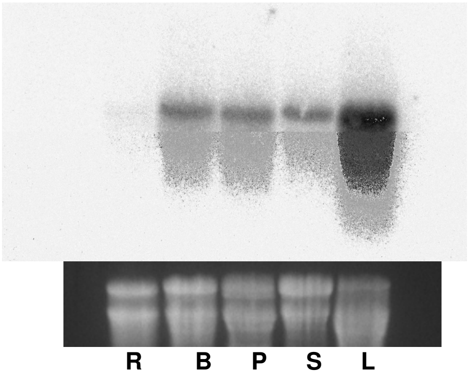

Figure 6. Tissue-specific expression of PR4 construct in transgenic cotton.
Highest expression is in leaves (L) and there is little or no expression in roots (R).
Lower levels of expression were seen in Boll coats (B), Petioles (P) and Stems (S)

10 Homozygous lines representing six independent transformation events as well as two different 35S lines and a non-transformed Coker 315 control were planted in the field at the Australian Cotton Research Institute, Myall Vale, Narrabri, NSW in the 2004-2005 growing season under OGTR License DIR049/2004. Preliminary GUS staining confirmed that the transgenic plants were expressing GUS and then leaf samples were collected at roughly two weekly intervals and stored at -80^{o}C. This material has only recently been transferred to Canberra and we are in the process of isolating RNA and assessing field expression of the small sub-unit promoter using Northern blotting so cannot yet provide definitive evidence for robustness of expression from this particular promoter.

In addition to the small sub-unit promoter some progress has been made with the Cab gene promoter, Elongation Factor-1 Alpha and Glyceraldehyde phosphate dehydrogenase and we have genomic clones stored away for later analysis and recovery of their promoters.

Conclusion

This project used both bioinformatics analysis of the abundance of ESTs and cDNA microarrays to identify cotton genes that have relatively robust expression throughout the cotton growing season. The identified targets were photosynthetic genes like the small subunit of Rubisco and Chlorophyll A/B binding protein genes, house keeping genes like elongation factor 1-alpha and tubulin and genes of primary metabolism like glyceraldehyde phosphate dehydrogenase (GPDH) as well as some unknown genes. Many of these genes turned out to belong to multigene families so it was important to verify expression patterns using gene specific probes to pick those members with the most robust expression. Initial experiments identified the small subunit promoter as a likely candidate and we progressed most with this promoter. The promoter was isolated and linked to a reporter gene and introduced into transgenic plants (both Arabidopsis, as a model and cotton). The promoter gave high levels of expression in the glasshouse and was expressed only in green tissues, as expected for a photosynthetic gene. Homozygous cotton plants were generated and approvals obtained to field trial them in Narrabri this last season. Leaf samples have been collected at fortnightly intervals and transferred back to Canberra where we are currently isolating RNA to look at expression of the introduced gene over the season from early vegetative to boll opening stages.

The project has identified the cotton small subunit promoter as capable of high level expression of transgenes (as high as or higher than the 35S promoter) although we are yet to find out how it is influenced under field conditions. A number of other candidates such as EF-1 Alpha and GPDH were identified as potentially highly expressed genes and some progress made towards isolating their promoters. These will provide useful resources for cotton researcher wishing to express genes at high levels in plants.

Outputs

Genetic engineering has shown great potential to improve the control of insect pests and weeds in cotton, however, the first commercial releases of INGARD and RR transgenic cotton failed to live up to expectations in Australia, because the expression of the introduced genes has not been uniform across the season, across the geographical spread of the industry, or even across farm units. Nearly all transgenic crops around the world rely on the 35S promoter from cauliflower mosaic virus (or very closely related viruses) to drive transgenes. It is only now becoming clear that this promoter is not as robust as laboratory and glasshouse studies have suggested and is influenced by as yet undefined environmental and physiological factors. In the short term we have relied on breeding to solve these problems of variable efficacy of transgenic cotton using the transgenes we already have in Australian cultivars, but in the longer term we must explore alternative promoters that, either on their own or in combination with the 35S promoter, will provide the industry transgene packages with robust expression throughout the season, irrespective of the source of the economically useful transgenes being used. This project aims to continue the process of isolating such genes and promoters from cotton so that a variety of promoters will be available for the next generation of transgenic cotton to be used by Australian growers. This will allow the use of more robust transgenes for pest, weed or disease resistance and ensure that the industry can continue to rely on the expression of these genes for crop protection without reverting back to unsustainable chemical controls.

Project Summary

This project has identified a number of potential genes and their promoters whose expression does not change drastically throughout the cotton growing season. The small subunit promoter was consistently identified and a 1829 bp fragment identified that gives strong expression in all photosynthetic tissues of both cotton and Arabidopsis and presumably in other dicot plants. Although US researchers have filed a patent on the 500 bp cotton small subunit promoter this is not in force in Australia and the current longer fragment appears to give significantly higher expression levels and could be considered an improvement on the US invention. The US patent was not filed in Australia and it is not clear that they will maintain their US patent given that it has not been used for any commercial products.

We also identified at least three other genes that could be useful targets for isolating promoters to drive high and robust expression in cotton but these would need to be developed further.

The project used novel genomic technologies to identify potential targets and probably represents the first use of this technology in cotton for promoter discovery. We do not believe this is patentable or requires correction of the IP register, but the information will shortly be submitted for publication once we have finished collecting all the data from the field trial. The EST collection will be lodged into public databases and the LSL clones have been included on our next generation cDNA microarray being used for other CSIRO projects.

Further research

Obviously not all the research envisaged for the project has been completed, some because of the premature departure of the post-doc working on the project. We had hope to have completed the analysis of the field trial by the time this report was submitted but have had to wait until technical staff in our Core Biotechnology project (CSP167C) had time to process the samples collected in the field. That work is currently underway and should be completed over the winter as time permits. The other genes and promoters identified in the project will have to be worked on as time is available but the clones isolated are safe in the freezer until further work can be carried out. Over the next year we would hope to be able to begin some isolation and characterisation of these promoters and identify the best targets for analysis in transgenic cotton. Given that the Cab gene, like RbcS, is also a photosynthetic gene it may be best to look at one of the house keeping genes like EF-1 alpha or the GPDH as a second type of promoter that is likely to be expressed in roots as well as green tissues so that we have a range of potential promoters available for cotton biotechnology. Over the next year we would also attempt to construct a plant expression vector containing the RbcS promoter and other components that are covered by CSIRO IP and make this available to other cotton researchers with an interest in transgene expression in cotton.

Publications

The research is currently being written up for publication, but is awaiting on more data to be generated from the field trial. We expect it to be published in an internationally recognised Journal such as Molecular Breeding or Molecular genetics and Genomics once it has been approved by CSIRO's internal editorial panel system.

Impact on the cotton industry

The impacts on the Industry are likely to be long-term as this project has developed or identified promoters that would be useful for various uses in biotechnology in the coming years. As new transgenes for pest or disease protection or weed control are developed, CSIRO and other Australian researchers will have a range of promoters that they could potentially use to drive their transgenes without relying on the 35S promoter that is already used to drive transgenes in commercial products. Products like Bollgard II/RR already contain half a dozen copies of the 35S promoter and this could lead to silencing and problems in the field as well as the issues about robustness of expression during the second half of the growing season. Promoters like the RbcS promoter will give researchers more flexibility and perhaps an easier route to commercialisation.

Part 4 – Final Report Executive Summary

The performance of the first generation of transgenic (INGARD) plants released in Australia since 1996 was disappointing at a commercial level, despite the considerable reduction in pesticide usage required to grow the crop. This was primarily because of the variable performance of the plants in insect control, either across farms, or at different times of the season. A decline in efficacy had previously been noted at the end of the season, but many cases of serious decline in insect control have been reported much earlier, requiring immediate spraying to control insect outbreaks. This decline in expression appears to be a consequence of the decrease in the activity of the 35S promoter driving the INGARD gene probably in response to either environmental or physiological influences on the plant. These problems cannot be corrected in the short-term using gene technology and we must rely on our breeders to select for individual plants that show more robust expression from the promoter driving INGARD. Monsanto has been able to resolve some of the problems with INGARD in their second generation Bollgard II cotton with higher levels of expression of the Cry2Ab gene. In the longer-term, for new biotech products, we can try to find better promoters that will express throughout the season or that at least are stronger during the period when the INGARD gene starts to decline. New gene constructs could then be developed that will either complement the existing genes or replace them and the same promoters could be used in conjunction with a number of other genes in the biotechnology pipeline. Using new genomic technologies we have identified a couple of possible promoters that might show the desired pattern of expression throughout the season, but they need to be fully evaluated in transgenic cotton plants. One promoter from a photosynthetic gene has been developed into gene constructs and introduced into both an easily engineered model plant and also into transgenic cotton, so that we can test its performance under field conditions. All the data is not yet to hand, but we hope that this, and other promoters to be analysed later, will give robust field expression in transgenic cotton and add to our toolkit of genes and pieces of genes from which we can develop more robust transgenic products for the cotton industry as well as provide biotechnologists with a greater selection of promoters to produce new biotech products.