Plant Biotechnology 23, 437–450 (2006)

Original Paper

Genomic approaches to the discovery of promoters for sustained expression in cotton (*Gossypium hirsutum* L.) under field conditions: expression analysis in transgenic cotton and Arabidopsis of a Rubisco small subunit promoter identified using EST sequence analysis and cDNA microarrays

B. H. R. Ranamalie Amarasinghe[1,a], Emmanuelle Faivre-Nitschke[1], Yingru Wu[1,4], Joshua A. Udall[2], Elizabeth S. Dennis[1], Greg Constable[3], Danny J. Llewellyn[1,*]

[1] CSIRO Plant Industry, Canberra, ACT, 2601, Australia; [2] Department of Ecology, Evolution and Organismal Biology, Bessey Hall, Iowa State University, Ames, IA, USA; [3] CSIRO Plant Industry, Narrabri, NSW 2390, Australia; [4] Food Standards Australia New Zealand, Canberra, ACT, 2600, Australia

* E-mail: Danny.Llewellyn@csiro.au Tel: +61-2-62465470 Fax: +61-2-62465000

Received August 24, 2006; accepted September 25, 2006 (Edited by T. Hashimoto)

Abstract Biotechnology requires robust and predictable expression of transgenes. Most commercial Genetically Modified (GM) crops contain the viral 35S promoter to drive insecticide and herbicide resistance genes. In cotton there have been reductions in efficacy of *Bacillus thuringiensis* toxin (Bt) expressing plants late in the season that have been attributed to reductions in promoter activity. We have used genomic approaches to identify cotton genes whose expression remains high during the season to find promoters that might better maintain expression of transgenes in the field. A cDNA library from young late season leaves was used to generate about 2000 ESTs. Clustering of ESTs was used to determine relative transcript abundance and identify the most highly-expressed genes. These were primarily photosynthetic and housekeeping genes and some metabolic genes. The ESTs were printed to a small cDNA microarray and probed with both early- and late-season leaf mRNAs. Absolute fluorescence levels were used to rank genes and confirm the EST abundance data. Candidate genes, including the small subunit of Rubisco (RbcS) were selected. An RbcS promoter (Genbank Accession DQ648074) was isolated and analysed in both Arabidopsis and cotton linked to a GUS reporter gene. Expression of the reporter gene was consistently high in green tissues throughout the life cycle of cotton in the glasshouse, and in the field. A number of other candidate promoters have been identified that may be useful in a variety of biotechnology applications.

Key words: Expressed sequence tags, gene expression profiling, promoter analysis.

Genetic engineering has shown great potential for improving the control of insect pests and weeds in cotton, however, the first commercial releases of transgenic insect-tolerant cottons in Australia (Ingard® cotton—called Bollgard® in other countries) have highlighted some deficiencies in their performance. Insect control efficacy has not been maintained across the season and appears to decline, particularly after squaring (flower bud formation) when the plants are committing resources to reproductive tissues (Fitt 2004). Variability of expression of the *Cry1Ac* gene has also been observed across the geographical spread of the plantings, or even across farm units. Part of the variability may be explained by environmental influences on insecticidal efficacy that do not involve changes in transcriptional or translational activity of the *Cry1Ac* transgene (Olsen et al. 2005b), but temporal changes in insecticidal efficacy during crop development appear to be, in part, due to a drop in the transcription of the introduced transgene towards the latter part of the growing season (Olsen et al. 2005a). The steady-state mRNA levels of both the *Cry1Ac* (Ingard®) gene and the selectable marker gene *Npt II*, both driven by the 35S promoter of Cauliflower Mosaic Virus (CaMV), decline

Abbreviations: AGRF, Australian Genome Research Facility; Cab, Chlorophyll a/b binding protein; CaMV, Cauliflower Mosaic Virus; FMV, Figwort Mosaic Virus; GAPDH, Glyceraldehyde 3-phosphate dehydrogenase; GEO, Gene Expression Omnibus; GUS, β-glucuronidase; MAST, Motif Alignment and Search Tool; NOS, Nopaline Synthase; NptII, Neomycin phosphotransferase; RbcS, Ribulose bisphosphate carboxylase, small subunit; X-gluc, 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid.

This article can be found at http://www.jspcmb.jp/

[a] Present address: Department of Plant and Soil Science, West Virginia University, Morgantown, WV, 26506, USA.

in later season plants (Olsen et al. 2005a), suggesting that this promoter is being influenced in a negative way by the developmental physiology of the plant. Nearly all transgenic crops around the world utilise the CaMV 35S promoter (Odell et al. 1985) (or similar promoters from 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 its function is influenced by as yet undefined physiological and perhaps environmental factors (Sunilkumar et al. 2002). In the short term we must rely on conventional breeding and selection to solve these problems of variable efficacy of transgenic cotton, but it will be useful, in the longer term, to identify other gene promoters that can drive strong expression of transgenes throughout the season. It is also important to have such promoters available for the next generation of transgenic cotton so that different traits can be stacked without relying on the same promoter so as to avoid transcriptional gene silencing induced by multiple copies of a single promoter such as the CaMV 35S promoter (see review by Fagard & Vaucheret 2000).

Promoter discovery for plant biotechnology has been largely empirical with most of the promoters deployed being isolated from plant pathogens such as viruses or *Agrobacterium* species (reviewed in Potenza et al. 2003). The *Agrobacterium* T-DNA provided a variety of constitutive promoters such as those from the octopine synthase and mannopine synthase genes and the useful but less active, at least in leaves, promoter from the nopaline synthase (NOS) gene. Viruses require high levels of transcription in a variety of tissues to ensure infection of a plant so it is not surprising that many strong promoters come from viruses. Most plant viruses have RNA, rather than DNA genomes so choices are limited. CaMV was one of the first plant viruses sequenced and provided two strong promoters responsible for the production of the 35S and 19S viral transcripts (Odell et al. 1985). Other viral promoters with similar characteristics to the CaMV 35S promoter have been, or are being, developed for use in plant biotechnology and these include the Banana Streak Virus promoters (Schenk et al. 2001), Cassava Vein Mosaic Virus promoter (Li et al. 2001; Verdaguer et al.1996), the Figwort Mosaic Virus (FMV) promoter (Sanger et al. 1990), the Mirabilis Mosaic Virus promoter (Dey and Maiti 1999), and the Subterranean Clover Stunt Virus promoters (Schünmann et al. 2003a, 2003b). Only the FMV promoter has so far appeared in any commercial transgenic crop plant, namely Bollgard II® cotton (Greenplate et al. 2003).

The second most common group of promoters for plant biotechnology have come from highly-expressed plant genes, such as those for seed storage proteins, photosynthetic proteins or housekeeping genes, all of whose mRNAs were easily cloned and characterised (reviewed in Potenza et al. 2003). Actin, ubiquitin and tubulin gene promoters have all been used in various plant species for expressing transgenes or selectable markers. As our sophistication in biotechnology improves, the need for more developmentally- or environmentally-regulated promoters has become evident and considerable effort is going into the discovery of specific tissue or biotic-, hormonal- or abiotic stress-responsive genes and promoters (reviewed in Potenza et al. 2003). The advent of EST and other genomic resources has changed the way researchers are approaching problems in biology and applied biotechnology and it is now possible to take more global approaches to finding more robust constitutive plant promoters.

In this paper we describe the EST and cDNA microarray-based approaches we undertook to identify potentially useful promoters in cotton and the detailed expression analysis of one such promoter, the Rubisco small subunit gene promoter, in both transgenic *Arabidopsis* and cotton plants.

Materials and methods

Synthesis of late season cotton leaf cDNA and construction of a leaf cDNA library

Total RNA was isolated from young, fully expanded, leaves of the transgenic Ingard cotton variety Sicala V3i collected 20 weeks after planting (i.e., during boll filling in the latter part of the growing season) from field-grown plants at Narrabri, NSW from 1.5–2.0 g of frozen leaf material (Wan and Wilkins 1994). Poly A^+ mRNA was prepared using an Oligotex mRNA kit (Qiagen, Cat. No. 70042), following the recommended protocol. 1.5 μg poly A^+ mRNA from the leaves was used for cDNA synthesis and a cDNA library was constructed using a Life Technologies' Superscript lambda system (Cat. No. 19643-014) following the manufacturer's instructions. The first strand cDNA was synthesized using a *Not* I primer-adapter, in the presence of 1 μCi [α-^{32}P] dCTP. The *Sal* I-adapted and *Not* I-digested cDNA was size-fractionated and the cDNA longer than 500 bp was directionally cloned in λZipLox *Not* I–*Sal* I arms (Life Technology Cat No. 15397-029). This primary library was comprised of 1.9×10^6 pfu with an average insert size of 0.9 kb.

EST Analysis and Design and Printing of cotton microarrays

Self-replicating plasmids were excised from the phage library according to the manufacturer's instructions and 2000 randomly selected clones were transferred to 96 well plates. Template preparation and sequencing was carried out at the AGRF (Australian Genome Research Facility, Brisbane Division) using the M13/pUC reverse primer. The sequences were assembled into clusters of related sequences with the TGI Clustering tools (TGICL), (http://www.tigr.org/tdb/tgi/software/), a software system for fast clustering of large EST

datasets. This package automates clustering and assembly of a large EST/mRNA dataset. The clustering is performed by a slightly modified version of NCBI's MEGABLAST, and the resulting clusters are then assembled using CAP3 assembly program as described in Udall et al. (2006). 1646 clones with reduced redundancy were selected from the original set for printing to the microarray.

The cDNA clones from the late-season leaf cDNA library were PCR amplified and printed onto CMT-GAPS coated microarray slides (Corning) using a VersArray ChipWriter Pro arrayer (Bio-Rad, Hercules, CA) as described in Wu et al. (2005). Post-printing slide processing was performed by baking the slides at 80°C for 3 hours as described in the manufacturer's technical manual.

Field sampling of cotton leaves for microarray experimental design

Leaf samples were collected from field-grown (Narrabri, NSW) cotton plants of the Sicala V-3BR variety that contains the Bollgard II and Roundup Ready transgenic insect resistance and herbicide tolerance traits. The youngest fully-expanded leaves from each plant were harvested at 8, 10, 14, 20, 22 weeks of age and about 100 randomly-selected plants sampled at each time point. Sampling was done on sunny days under similar conditions between 10–11 am to account for any diurnal differences in gene expression. Leaf samples were stored at −80°C until used to isolate RNA (Wan and Wilkins 1994). RNAs from the same plant variety from different fields were sampled on the same dates and served as a biological replicate. Developmentally, the cotton plants in the two fields were roughly at the same stage having been planted on the same day.

RNAs from 8-week old plants were compared to RNAs from 10-, 14-, 20- and 22-week old plants on separate arrays each with a dye swap in a common reference experimental design. The 20 and 22 weeks hybridisations each had a biological replicate (each with a dye swap) as well as a second independent labelling and hybridisation (with dye swap) with one of the 22 week RNAs. This generated 14 sets of two channel fluorescence data, although one of the 22-week slides was of low quality and had to be discarded. tRMA (Wilson et al. 2003) was used to normalise the raw data as described in Wilson et al. (2005) (so that different array experiments could be compared) and to identify the statistically significant differentially expressed genes.

Microarray probe preparation and slide hybridisation

Preparation of polyA$^+$ RNA, probe labelling with Cy3-dUTP and Cy5-dUTP (Amersham Pharmacia Biotech) slide hybridisation and washing were as described in Wu et al. (2005). Microarray images were captured using a GenePix 4000A microarray scanner (Axon Instruments, Union, CA, USA).

Scanned microarray images were analysed using the GenePix Pro program (Axon Instruments, Union, CA, USA). Grids were predefined and manually adjusted to ensure optimal spot recognition and bad spots were flagged (eg. dust contamination etc.). Spots were quantified using the GenePix's fixed circle method, and medians of the fluorescence intensity of the red and green channels were used to generate the ratio of the two channels and to calculate the absolute fluorescence values.

The statistical analysis of the microarray data was carried out using tRMA (tools for R Microarray Analysis, Version), a suite of statistical functions written in R code as described in Wu et al. (2006). For a typical microarray comparison in this study that consisted of 4 hybridizations (two biological replicates and two dye swaps), only genes occurring in at least 3 or more of the 4 replications were counted as differentially expressed. All the microarray data and layout of the clones on the slides is available at http://www.pi.csiro.au/gena/ and in The Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) as Series Accession No: GSE4939 with Platform Accession No: GPL3823 describing the array.

Cloning and Sequencing

All DNA manipulations were performed using standard methods (Sambrook et al. 1989) and manufacturer's recommendations where appropriate. Sequencing was carried out using Big Dye Terminator chemistry (Applied Biosystems, Foster City, CA, USA). DNA sequences were analyzed and manipulated with the Wisconsin GCG Package v 9.1 (Genetics Computer Group, Madison, WI, USA, 1997) and individual BLAST searches were conducted via the National Center for Biotechnology web site (http://www.ncbi.nlm.nih.gov).

Small Subunit promoter cloning and T-DNA Vector Construction

280,000 plaques were screened from an amplified *G. hirsutum* cv. Deltapine16 partial *Sau*3AI genomic library in λEMBL4 (Stratagene) (J. Norman and R. Chapple, CSIRO Plant Industry, Canberra) on duplicate lifts with either the RbcS 5′ promoter region-specific probe generated from genomic DNA using the primers OL5 (GCGAATTCCGCTCATGTTAACAATTAATTC) and OL6 (GCGGATCCCATTGCTATTACTGCTTACTAG) or a coding region probe excised from the RbcS cDNA clone LSL0087D07. Eight plaques positive for both probes were selected and purified by two more rounds of hybridization and lambda DNA was prepared as recommended using a Lambda Mini kit (Qiagen, Hilden, Germany). All DNAs had identical restriction patterns with *Bam* HI or *Eco* RI so only one was sub-cloned and sequenced. A 3.8kb *Bam* HI fragment hybridizing to the RbcS 5′ specific probe was subcloned into pBluescript (Stratagene) and approximately 2 kb upstream of the start codon was sequenced using a primer-walking strategy from the 5′ end of the coding region.

RbcS promoter fragments were amplified from the 3.8 kb *Bam* HI subclone using a 3′ primer adjacent to the start codon (OL34:GC**TCTAGA**TGCTATTACTGCTTACTAGTAC) and a 5′ primer at various distances upstream to generate fragments of 397, 723, 925 and 1827 bp. The 5′ primers (OL35:CCC-**AAGCT**TGACCAAGCAAACAAGGTATGG; OL36:CCC**AAGCTT**GCTTTCAATGTTGCGGGGTC; OL37:CCC**AAGCTT**CTCACATTACTGGGTCCTGTTCG; OL38:CCC**AAGCT**-**T**CGGTGATAGAAAAAGGCAAGG) each contained a *Hind* III restriction site (in bold type) and the OL34 primer an *Xba* I restriction site (in bold type) and were first cloned in pGEM-T and sequenced to ensure they contained no PCR-generated

errors. Promoter fragments were then cloned into the promoterless-GUS T-DNA vector pIG121Hm (Ohta et al. 1990) digested with the *Hind* III and *Xba* I. pIG121Hm contained the plant-selectable markers for resistance to kanamycin and hygromycin.

Arabidopsis transformation

Transformation with the various RbcS promoter GUS constructs in *Agrobacterium tumefaciens* strain AGL1 were carried out by the floral dip method (Clough and Bent 1998). Transgenic T0 plants were identified by plating seeds from the dipped plants on MS medium containing 100 μg/ml kanamycin sulphate and green kanamycin-resistant plants were transferred to soil in the glasshouse to generate T_1 seed. A dozen T_1 plants were grown to produce T_2 seed which was subsequently screened on plates containing kanamycin to identify homozygous lines that were used for quantitative GUS assays.

Cotton transformation

Agrobacterium-mediated transformation of *G. hirsutum* L. cv. Coker 315 was performed using cotyledon segments as detailed in Murray et al. (1999). Healthy plantlets were transferred to soil in a growth cabinet (28°C 16 h light; 22°C 8 h dark), humidified with plastic covers, and moved to a glasshouse when established.

Cotton Plant growth conditions

Plants were grown in potting mix and allowed to self pollinate in conditions of 28°C day temperature (16 h) and 20°C night (8 h). Seed was ginned and acid delinted with concentrated sulphuric acid prior to planting. Transformants were confirmed on young plants by histochemical staining for GUS on young leaf tissue before transfer to the glasshouse and then in each generation thereafter to follow the inheritance and segregation of the introduced reporter gene.

Field Growth of Cotton

All RbcS and 35SGUS plants were grown under License (DIR049/2003) from the Office of the Gene Technology Regulator. The cotton was grown on raised beds 1 m apart at a density of about 10–13 plants per metre for large scale plantings (3 plants per metre for small scale plantings of the GUS plants). Irrigation, fertiliser application and pest and weed control were as dictated by standard industry practice.

Histochemical GUS assays

For histochemical GUS assays whole *Arabidopsis* seedlings or pieces of cotton leaves, squares, bolls etc., were vacuum infiltrated for 10 min at 25 mm Hg at 37°C in GUS staining buffer (0.5 mM $K_2Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6 \cdot 3H_2O$, 0.1 mM phosphate pH 7.0, 10 mM EDTA, 1.5 g L^{-1} 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid, cyclohexylammonium salt (X-gluc; Diagnostic Chemicals Limited, Oxford, CT, USA), 0.5% dimethyl sulfoxide) and stained overnight or as indicated. Tissues were dehydrated and rehydrated via a graded ethanol series to a final 10 mM phosphate buffer pH 7.0 for bright-field microscopy.

Flurometric GUS assays

Leaf or other tissues samples were ground in Buffer A as described in Breyne et al. (1993). GUS fluorometric assays were carried out as described using 4-methylumbelliferyl β-D-glucuronide substrate (MUG; Diagnostic Chemicals Limited, Oxford, CT, USA). The rate of 4-methylumbelliferone (MU) production was measured using a Fluoroscan II (Labsystems, Vantaa, Finland) and expressed as pmol MU per minute per microgram of total soluble protein ($min^{-1}\,\mu g^{-1}$), using MU sodium salt (ICN Biochemicals Inc., Aurora, OH, USA) to generate a standard curve for the conditions used. Protein concentrations of extracts were determined using Bio-Rad protein reagent (Bio-Rad, Hercules, CA, USA) according to the manufacturer's instructions. The enzyme reactions and protein determinations on each sample were carried out in triplicate and averaged.

DNA isolation and Southern blot analysis

Southern blot hybridisation was performed to determine T-DNA insertion number and segregation of insertion events. Genomic DNA was extracted from plants, digested (20 μg) with restriction enzymes, electrophoresed in a 0.8% agarose gel, blotted to Hybond-N^+ nylon membranes (Amersham, Buckinghamshire, England) and hybridised with a GUS probe made from an *Eco* RI/*Nco* I fragment of approximately 1800 bp containing the whole β-glucuronidase-coding region labelled with α^{32}P-dCTP (Amersham, Buckinghamshire, England) as described in Townsend and Llewellyn (2002).

Northern blot analysis

Total RNA was extracted as described above and 15 μg was electrophoresed in formaldehyde-agarose, blotted to Hybond-N nylon membrane and probed with GUS coding region probes as described in Townsend and Llewellyn (2002). The hybridised filters were analysed using a Phosphorimager (Molecular Dynamics) and ImageQuant version 3.3 software (Molecular Dynamics). Riboprobes were synthesised with the Promega (Madison, WI, USA) Riboprobe—System SP6 kit (Cat. No. P1420) as described by the manufacturer using linearised plasmids containing the appropriate ESTs. Gene-specific riboprobes were made by linearising the plasmids at restriction sites near the end of the coding region so that only the 3′ untranslated region was transcribed.

Results

A Late Season Leaf cDNA library and preliminary DNA sequence analysis

A high-quality directional cDNA library was prepared from leaf RNA extracted from late-season field-grown transgenic Ingard cotton plants (containing the *35S-Cry1Ac* insecticidal gene and the *35S-NptII* selectable marker gene). Over two thousand cDNA clones were randomly selected and sequenced from their 5′ ends. Approximately 1810 high-quality sequences were recovered and clustered into 1340 contigs as described in the Materials and methods. BLAST searching was used to identify the most closely-related entry in the

SwissProt protein database, or when there was no significant hit, to the Genbank Non-Redundant nucleotide database. These EST sequences are available from Genbank (Accessions DV848634 to DV850414) and the singletons and contigs are available as the GH_LSL cDNA library clones from the cotton genomics database (Udall et al., 2006) at the Arizona Genomics Institute using the PAVE sequence browser (http://www.agcol.arizona.edu/pave/cotton/).

To define the most highly-expressed genes at a 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 grouped 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 higher order groupings, and the corresponding number of total ESTs in each, are shown in Table 1. The most abundantly-expressed gene in late season leaves appeared 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, ranking second and third most abundant. House keeping genes such as elongation factor 1, and tubulin were also highly represented. Actin was present, but was less abundant than 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. Several of the higher order groupings were composed of a large number of contigs, indicating that they were probably not from single genes, but from different members of multigene families. The neomycin phosphotransferase (*Npt II*) gene, used as a selectable marker in the commercial transgenic cotton variety from which the library was made, appears in the top twenty most abundant assemblies. These plants were homozygous for a single-copy insertion of the *NptII* gene driven by the CaMV 35S promoter.

Table 1. Assembly of LSL ESTs into their most abundant contigs. Numbers of ESTs indicate the total number of ESTs in all contigs with an identical best match in a Megablast or Blast N search. The small subunit of Rubisco (RbcS) and kanamycin (Npt II) resistance genes are indicated in bold type.

# ESTs	# Contigs[1]	Probable Identity
253	1	26S ribosomal RNA gene
69	15	**Rubisco Small subunit (RbcS)**
23	9	Chlorophyll a/b binding protein (Cab)
11	9	Elongation Factor-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
7	2	germin-like protein
6	4	Fructose bisphosphate aldolase
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	Elongation Factor-2

[1] six different contigs with 3 or more ESTs, but no significant hit with Megablast or BlastN, not included.

Analysis of the temporal expression patterns of cotton leaf ESTs using cDNA microarrays

To get a more complete picture of the temporal differences in expression of the late season ESTs, 1646 clones were printed to glass slides and probed with labelled cDNAs from leaves taken at different times during development of cotton from early vegetative to fruit development and maturation stages. 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 transgenic Bollgard II/Roundup Ready cotton plants growing in large commercial field plantings rather than from plants grown in a glasshouse. RNA isolated from two biological leaf replicates of young pre-flowering, vegetative plants (8 weeks) was compared on the leaf cDNA microarray to leaf RNA isolated from relatively-mature plants (20- and 22-weeks old) when the plants were carrying their peak boll (fruit) load, or were starting to open their bolls, respectively. Surprisingly, few genes were identified as being differentially expressed in leaves at the later stages of growth compared to similar leaves from younger vegetative plants (40/1646 at 20 weeks and only 10/1646 at 22 weeks, the majority being down regulated at the later stages). Tables 2 and 3 show the genes that were up- or down-regulated in cotton leaves aged 20 and 22 weeks, respectively with the false discovery rate controlled at 0.001. At 20 weeks only two clones were up-regulated compared to their expression level at 8 weeks (Table 2). These two ESTs were homologous to a chlorophyll A/B binding protein and were 89% identical at the nucleotide level, so were from different genes. At 22 weeks all 10 differentially expressed genes were down-regulated (Table 3) including the 35S-NptII gene, consistent with previous reports of loss of expression during the season. The EST, LSL001DO9, homologous to the tannin biosynthetic gene leucoanthocyanidin reductase, showed the highest decrease in expression, 10-fold compared to its level at 8 weeks of age. This gene was also significantly down-regulated at 20 weeks (Table 2).

Since there were no genes that consistently increased

Table 2. Genes differentially expressed in leaves at 20 weeks compared to 8 weeks in field grown cotton plants. The dotted line separates up- from down-regulated genes.

Clone Name	Average $\log_2$ ratios (Cy3/Cy5 and Cy5/Cy3)	Std Dev	Back-transformed Ratio	Best Match in SwissProt[1]
LSL030E06	0.872	0.2287	1.831	Chlorophyll a/b protein (2E-107)
LSL007E01	0.711	0.0542	1.637	Chlorophyll a/b protein (1E-100)
LSL010D04	−0.699	0.1829	0.615	Rubisco Activase (4E-102)
LSL026B02	−0.745	0.3151	0.596	Anther-specific proline rich protein (2E-63)
LSL021E10	−0.760	0.1371	0.590	Glycine decarboxylase (6E-54)
LSL024B11	−0.786	0.1464	0.579	Monodehydroascorbate reuctase (2E-103)
LSL026F02	−0.810	0.2092	0.570	Vacuolar pyrophosphatase (1E-102)
LSL021D01	−0.811	0.2250	0.569	ATP-citrate synthase (3E-100)
LSL023A02	−0.815	0.0939	0.568	Chalcone synthase 1 (6E-98)
LSL028F07	−0.828	0.0951	0.562	S-adenosylmethionine synthetase 1 (1E-111)
LSL009A04	−0.837	0.1755	0.559	Glutathione peroxidase (1E-79)
LSL023F05	−0.841	0.2260	0.558	Acetyl-CoA carboxylase 2 (7E-99)
LSL010G12	−0.851	0.1420	0.554	Heat shock protein (1E-82)
LSL009G12	−0.866	0.0867	0.548	Peroxisomal-coenzyme A synthetase (1E-52)
LSL007F03	−0.870	0.1583	0.547	Glyceraldehyde-3-phosphate dehydrogenase (2E-100)
LSL001H10	−0.881	0.3126	0.542	S-adenosylmethionine synthetase 1 (2E-92)
LSL021F11	−0.883	0.4733	0.542	Flavonoid 3′,5′-hydroxylase 2 (4E-122)
LSL001G08	−0.890	0.1235	0.539	Putative peroxisomal-coenzyme A synthetase (3E-26)
LSL024A05	−0.904	0.1006	0.534	Glutathione peroxidase (7E-82)
LSL009H11	−0.915	0.2447	0.530	Chalcone-flavonone isomerase (4E-51)
LSL006E01	−0.923	0.1202	0.527	Glyceraldehyde-3-phosphate dehydrogenase (7E-66)
LSL022E10	−0.948	0.1092	0.518	Ubiquitin (6E-108)
LSL001D01	−0.972	0.1220	0.509	RuBisCO activase (1E-77)
LSL006B08	−0.979	0.1676	0.507	5-methyltetrahydropteroyltriglutamate-Homocysteine methyltransferase (5E-86)
LSL010H03	−1.015	0.0485	0.494	Heat shock cognate 70 kDa protein 1 (3E-64)
LSL001A12	−1.016	0.1358	0.494	Putative cell wall protein precursor (5E-08)
LSL003D06	−1.025	0.1722	0.491	Chalcone synthase 1 (1E-108)
LSL022D11	−1.025	0.2212	0.491	RuBisCO activase (1E-113)
LSL024A10	−1.046	0.2502	0.484	Protein serine/threonine receptor kinase (1E-95)
LSL022C09	−1.070	0.1296	0.476	Early light-induced protein, chloroplast precursor (ELIP) (8E-16)
LSL007C04	−1.081	0.1803	0.472	Probable pyridoxin biosynthesis protein ER1 (ethylene-inducible) (6E-99)
LSL031H07	−1.091	0.2734	0.469	Probable pyridoxin biosynthesis protein ER1 (ethylene-inducible) (2E-46)
LSL008F10	−1.120	0.2134	0.460	Alanine aminotransferase 2 (3E-74)
LSL005G02	−1.233	0.1584	0.425	Dehydrin (7E-08)
LSL007C10	−1.246	0.1897	0.421	Zinc finger protein (3E-25)
LSL023F08	−1.266	0.2956	0.415	no match (Novel gene)
LSL007E05	−1.269	0.1144	0.414	Cell wall protein precursor (3E-09)
LSL025E09	−1.328	0.0098	0.398	Heat shock protein (5E-101)
LSL004A12	−1.679	0.1842	0.312	Catalase isozyme 2 (2E-113)
LSL001D09	−2.430	0.1168	0.185	Leucoanthocyanidin reductase (1E-124)

[1] Best match by BlastX (E-values for alignment with database sequence)

Table 3. Genes differentially expressed in cotton leaves at 22 weeks compared to 8 weeks in field grown plants.

Clone Name	Average Log_2 ratio (Cy3/Cy5 and Cy5/Cy3)	Std Dev	Back-transformed Ratio	Best Match in SwissProt[1]
LSL009A08	−1.108	0.6031	0.463	NptII (kanamycin resistance gene) (6E-114)
LSL026B02	−1.195	0.1064	0.436	Anther-specific proline rich protein (2E-63)
LSL004H10	−1.219	0.3300	0.429	NptII (kanamycin resistance gene) (7E-108)
LSL003D06	−1.239	0.1836	0.423	Chalcone synthase (1E-108)
LSL024A10	−1.287	0.3160	0.409	Protein serine/threonine receptor kinase (1E-95)
LSL028C09	−1.311	0.3260	0.403	Glutathione S-transferase
LSL005G02	−1.318	0.1913	0.401	Dehydrin (7E-08)
LSL009H11	−1.416	0.1424	0.374	Chalcone—flavonone isomerase (4E-51)
LSL022C09	−1.566	0.2666	0.337	Early light-induced protein (ELIP) (8E-16)
LSL001D09	−2.955	0.2067	0.129	Leucoanthocyanidin reductase (1E-124)

[1] Best match by BlastX (E-values for alignment with database sequence)

Table 4. Most highly expressed ESTs averaged over all developmental ages. Values represent the average of the $\log_2$ transformed median fluorescence values in the red and green channels of each spot on the array averaged for that spot over all of the slides in the developmental comparison (covering leaves from 8 to 22 weeks of age). 26S and 18S ribosomal RNAs and any clones for which we do not have reliable sequence information have been removed from the list.

Clone Name	Average Log Intensity[1]	SE	Average Log Ratios[2]	SE	Best Match from BlastX	E value for BlastX
LSL008D07	14.4618	0.117	−0.36	0.10	RbcS	9E-98
LSL011E08	14.4266	0.089	−0.31	0.11	RbcS	1E-86
LSL010F03	14.3326	0.082	−0.30	0.10	RbcS	1E-95
LSL010F04	14.2634	0.087	0.21	0.16	2-Oxoglutarate-Fe(II) oxygenase	2E-26
LSL008B03	14.2311	0.050	0.29	0.15	No match	—
LSL010C09	14.1707	0.050	0.29	0.15	putative senescence-associated protein	1E-68
LSL009E07	14.1252	0.151	−0.36	0.12	RbcS	3E-90
LSL006G03	14.0969	0.139	−0.32	0.10	RbcS	5E-97
LSL011H06	14.0532	0.112	−0.38	0.11	RbcS	7E-97
LSL008B11	14.0334	0.078	−0.43	0.09	RbcS	7E-67
LSL011B10	13.9481	0.085	−0.25	0.08	At hypotheical protein	2E-60
LSL022E07	13.9036	0.063	0.41	0.15	No match (short sequence)	—
LSL007B09	13.8800	0.088	−0.31	0.10	RbcS	2E-87
LSL008A03	13.8646	0.094	−0.37	0.11	RbcS	4E-96
LSL009H10	13.7796	0.076	−0.19	0.09	RbcS	2E-45
LSL011B01	13.7541	0.076	0.37	0.14	No match	—
LSL024E02	13.7177	0.066	−0.12	0.09	No match	—
LSL005C03	13.7143	0.120	−0.32	0.10	RbcS	1E-83
LSL025G03	13.7028	0.109	−0.15	0.11	RbcS	3E-49
LSL024G05	13.6941	0.068	0.01	0.09	PS I Reaction Centre protein	7E-22
LSL005A06	13.6472	0.081	−0.35	0.10	RbcS	5E-96
LSL025H01	13.5621	0.076	−0.14	0.11	s-adenosyl methionine decarboxylase	1E-47
LSL025H03	13.5576	0.078	0.01	0.09	No match	—
LSL006G09	13.5557	0.089	−0.12	0.11	sedoheptulose-1,7-bisphosphatase	4E-88
LSL024D06	13.5265	0.094	0.01	0.10	No match	—
LSL021C01	13.4724	0.063	0.25	0.14	putative phosphatase	2E-90
LSL005D12	13.4559	0.084	−0.34	0.10	RbcS	5E-93
LSL022H05	13.3964	0.134	0.20	0.11	Chlorophyll a/b binding protein	1E-136
LSL007F11	13.3408	0.153	−0.25	0.10	RbcS	9E-61
LSL025E01	13.2882	0.105	−0.02	0.12	Glucosyltransferase	2E-91
LSL007F12	13.2579	0.148	0.31	0.13	Chlorophyll a/b binding protein	1E-128
LSL024C08	13.2516	0.267	−0.04	0.11	No match	—
LSL025D11	13.2397	0.327	−0.08	0.10	Hypothetical protein	1E-111
LSL006G08	13.2353	0.081	−0.10	0.11	40S ribosomal protein S7	4E-86
LSL006D10	13.1769	0.119	0.06	0.08	No match	—
LSL006E10	13.1550	0.092	0.01	0.10	No match	—
LSL001D01	13.1157	0.158	−0.67	0.13	RbcS	1E-76
LSL011E03	13.0881	0.091	0.01	0.09	PSI reaction centre subunit VI	4E-29
LSL024H11	13.0728	0.121	−0.02	0.09	Central motor kinesin 1	5E-31
LSL025A04	13.0715	0.108	−0.05	0.10	T-complex protein 1 zeta subunit(Animal)	6E-70
LSL004C07	13.0522	0.153	−0.12	0.11	No match	—
LSL003A06	13.0494	0.071	−0.37	0.10	RbcS	6E-93
LSL004H10	13.0333	0.094	−0.58	0.20	NptII	7E-108
LSL025C01	13.0175	0.077	0.02	0.10	No match	1E-30
LSL006D03	13.0102	0.167	−0.07	0.11	putative G-box binding protein	1E-31
LSL006D11	13.0043	0.194	−0.05	0.10	No match	—
LSL008H04	13.0007	0.108	0.06	0.09	Protein translation factor SUI1 homolog	3E-20
LSL011D11	13.0002	0.134	0.21	0.12	Chlorophyll a/b binding protein	6E-111

[1] $\log_2(R*G)/2$ of each spot averaged for all slides in the series

[2] $\log_2(R/G$ or $G/R)$ for each spot to give an early sample versus later sample fluorescence ratio averaged over all slides in the series

in expression from early to late season we focussed on those genes that had relatively uniform high expression throughout the life cycle of cotton. Many of the clones with high log-intensity values also had log ratios of early to late expression (averaged across all comparisons) that were close to zero (Table 4) indicating that they were also uniformly expressed in different stages of leaf development. The log 2 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. Table 4 summarizes the ESTs with highest average log 2 transformed fluorescence value (i.e., 13–15) over 8–22 weeks of age determined from

the 13 good slides. Excluding a few ESTs with very short or poor sequence and hence no convincing database matches, the top ten ESTs were all 26S ribosomal RNA genes (not shown in the Table), consistent with their EST abundance. The next most-abundant class of ESTs corresponded to the photosynthetic carbon fixation genes of the Rubisco small subunit family, although with a broad spread of average fluorescence levels suggesting that different gene family members had different expression levels.

The *NptII* selectable marker gene was represented by two ESTs on the array (LSL004H10 and LSL009A08) and these had average fluorescence values of 13.03 and 12.48, respectively, so we expected genes with fluorescence values of 13 and over to represent a high average level of expression given there are two copies of the *35SNptII* transgene per haploid genome in the plant material from which the RNA was isolated (one introduced with the *Cry1Ac* insecticidal gene and one with the glyphosate herbicide resistance trait). Over 440 ESTs had expression levels higher than 11. House keeping genes like alpha and beta tubulin, ubiquitin and translation elongation factors tended to be in the mid range with fluorescence values between 12 and 13. All the negative control genes had average fluorescence values of less than 7.0 (not shown) so we expect that genes with fluorescence values of 7–8 or less are very low abundance and of no interest.

Verification of the microarray data by northern blotting

To examine the expression levels of some of these genes at various stages of growth, and to confirm our microarray data, northern blot analysis was performed on total leaf RNA isolated from field-grown Bollgard II/Roundup Ready cotton plants at the same growth stages as those used for the microarray analyses. Coding region riboprobes complementary to the EST LSL008D07 (the RbcS EST with the highest average fluorescence value) revealed that the small subunit of Rubisco family of genes were very highly expressed throughout the season (Figure 1A; with much higher hybridisation signals than a similarly-labelled ubiquitin probe from EST LSL028G08, data not shown). A gene-specific probe from the 3′ end of LSL008D07 gave a much lower signal on a northern blot hybridisation (Figure 1A), but indicated that expression in leaves was relatively steady from early vegetative to the boll-setting phase. Sequence-specific riboprobes made against the EST LSL030E06 (Cab), noted as being about two fold up-regulated later in the season, or the EST LSL001D09 (leucoanthocyanidin reductase), highly down-regulated in late season leaves, revealed expression changes consistent with those seen on the microarrays. The Chlorophyll a/b-binding protein gene was up-regulated at

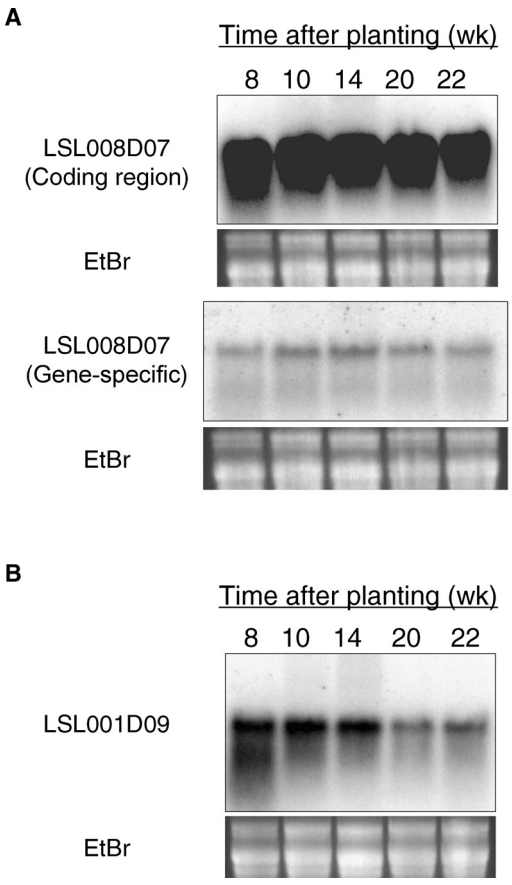

Figure 1. Confirmation by northern blot analysis of the expression profiles in leaf tissues over the life cycle of cotton of some selected ESTs identified from microarray experiments. (A) Rubisco small subunit gene represented by EST LSL008D07 (Upper panel: coding-region probe using the whole EST as a riboprobe, Lower panel: hybridised to a 3′ end gene-specific riboprobe containing mainly the 3′ UTR). Panel below each blot are the ethidium bromide stained gel before transfer to the nylon membrane. Times above lanes lane are weeks post planting. (B) EST LSL001D09 corresponding to a leucoanthocyanidin reductase gene that was significantly down regulated according to the microarray experiments. Blot probed with a 3′ end-specific riboprobe.

20 and 22 weeks compared to 8 weeks (not shown), while the leucoanthocyanidin reductase gene was significantly down-regulated in the leaves from older plants (Figure 1B). Other genes verified by northern blot hybridisation analyses, (GAPDH, Elongation Factor-1a, and ubiquitin) were expressed approximately equally early and late in the season (not shown) indicating that the hybridisation data collected from the microarrays was a good representation of the steady-state mRNA levels in field grown plants.

Isolation of the promoter region of the Rubisco small subunit gene of cotton

Evidence from both EST abundance and strength of hybridisation on the microarray suggested that the Rubisco small subunit gene was a good candidate from which to isolate a strong seasonally-insensitive promoter.

The EST LSL008D07 which had the highest expression throughout the vegetative and reproductive stages (Table 4) was identical in its sequence to the cotton RbcS small subunit gene sequence available in the Genbank database (X54091). We used the LSL008D07 EST and an X54091 promoter-specific probe (Materials and Methods) to screen a cotton (cv. Deltapine 16) genomic library to isolate the corresponding RbcS gene (*GhRbcS*) along with a 1829 bp 5′ promoter region. This promoter was identical over the first 521 bp to that of X54091, but extended a further 1308 bp in the 5′ direction. The gene had three exons and three introns with sequence identical to X54091 despite being from a different cultivar. This is consistent with the low level of DNA polymorphism reported within cotton cultivars (Rungis et al. 2005; Small et al. 1999). The longer promoter sequence has been lodged in Genbank as Accession DQ648074. The cotton Rubisco promoter sequence was compared to *RbcS* gene sequences from other plants (coffee, chrysanthemum, tomato, *Arabidopsis*, maize, and rice) by MAST (Motif Alignment and Search Tool) (available at http://meme.sdsc.edu; Bailey and Gribskov, 1998). The dicot promoters were the most similar with the highest similarity of the cotton promoter being to the chrysanthemum RbcS upstream region (about 18% identity). Four consensus motifs were identified in the cotton promoter shown schematically in Figure 2A and these were common to many of the dicot RbcS promoters. An additional GC-rich motif #1 (GCCGGGCTGCCCGGCCGCG**GCCGCC**GGCG) appeared to be confined to the monocot RbcS promoters and was not present in the cotton or other dicot promoters. It contains the core GCC-box (bold) found in many pathogen-responsive genes and has been shown to function as an ethylene-responsive element (Brown et al. 2003). The most-prevalent motif in all the plant RbcS promoters, consensus Motif #3 (GAAATATATGC**ATTTTTA**T**TTTTTC**TCTTGTTGT TTTTTGCAAACAA), occurred multiple times in both orientations in the upstream regions of all the dicot promoters, including the cotton RbcS (Figure 2A), but not in the chrysanthemum promoter. It is the potential binding site for at least two transcription factors and contains a binding site (GAAAAA) for GT-1 (Manzara et al. 1991) on the reverse strand and a potential site (ATTTTTA) for the soybean embryo factor 4 protein (SEF-4) on the positive strand (both in bold type). The GT-1 binding site is common in the promoters of many light regulated genes. Three other consensus motifs #2, #4, and #5, being **GTGGTCATTAAGTATG**TAAT-GTCATG**AGCCACA**GGATCCAATGGC, CTG**CCAC-GTGG**C, and GGTGGTCAAT**GATAAG**G, respectively, were found in many of the dicot RbcS promoters, including the cotton promoter. In the majority of the promoters there was a consistent arrangement of closely-spaced motifs (in the order: Motif #4 + Motif #5 + Motif #2) just upstream from the CAAT and TATA boxes. The 5′ end (bold type) of the consensus Motif #2 has some similarity to the Box II sequence (another binding site for GT-1) common in light regulated promoters (Manzara et al. 1991) and also contains a conserved motif of unknown function, SORLIP-1 (Sequences over-represented in light-induced promoters), (bold type) seen in the promoters of a many light responsive genes (Hudson & Quail, 2003). The putative Box II sequence in the Motif #2 of the cotton promoter is well conserved relative to the other dicot RbcS promoters, but the region with the SORLIP-1 sequence has diverged from the consensus, so is unlikely to be a

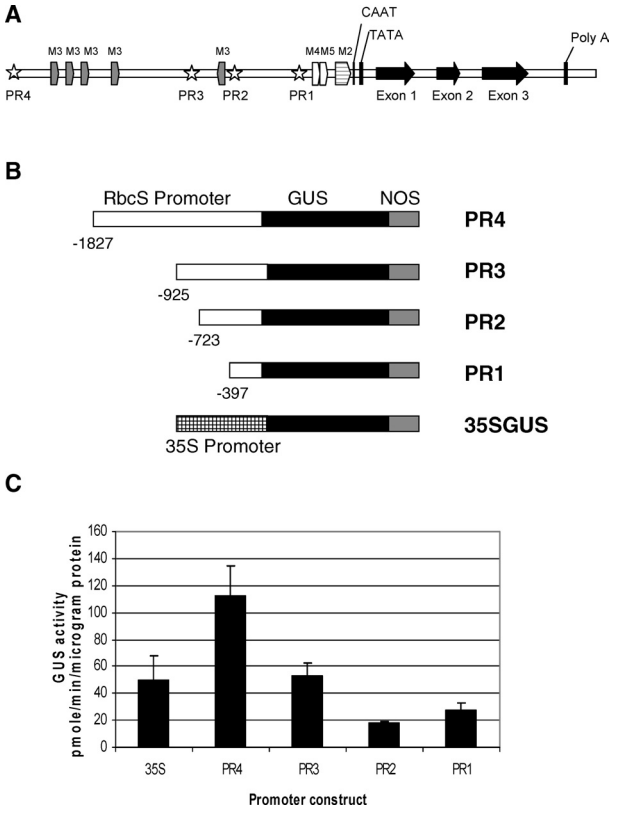

Figure 2. Structural and functional analysis of the cotton small subunit promoter. (A) Schematic of the cotton RbcS promoter (Genbank Accession DQ648074) and gene structure and the consensus motifs identified by MEME. Exons are shown as black block arrows. Putative TATA, CAAT and polyadenylation signals are indicated as are the various conserved sequence motifs (M2-M5) shared with other dicot RbcS promoters (coffee, *Arabidopsis*, tomato, and chrysanthemum). End-points of the various 5′ deletion constructs are marked with stars. (B) Schematic of the promoter GUS constructs introduced into *Arabidopsis*. The various 5′ deletions of the RbcS promoter (open boxes) are indicated with the endpoints relative to the start of translation. The GUS coding region is indicated by black boxes and the terminator of nopaline synthase (NOS) by grey boxes. The 35S promoter is hatched. (C) Quantitative analysis using a MUG substrate of GUS expression levels in six independent transgenic *Arabidopsis* lines each transformed with either the full-length or various 5′ deleted cotton RbcS promoter GUS constructs compared to similarly transformed 35S-GUS lines.

functional transcription factor binding site in this particular gene. Motif #4 contains the G-Box element (bold type) known to be the binding site for the GBF family of bZIP transcription factors (Donald and Cashmore 1990), while Motif #5 contains the classical I-box or GATA-box element (bold type), the binding site for the transcription factor GA-1 (Donald and Cashmore 1990), which together are known to confer light responsiveness to RbcS promoters (reviewed in Argüello-Astorga and Herrrea-Estrella 1998). These binding sites are both highly conserved in the cotton promoter.

Functionality of the Rubisco promoter in Arabidopsis

Because RbcS is encoded by a small gene family it was important to verify that the promoter we isolated was functional and not from a pseudo gene. The full-length RbcS promoter and three 5′ deletion constructs removing some of the important conserved motifs of the GhRbcS promoter (promoter constructs PR4, PR3, PR2 and PR1, which were 1827, 925, 723 and 397 bp, respectively, in length upstream of the RbcS translation initiation site) as well as a CaMV 35S promoter, were fused to the GUS reporter gene (Figure 2B) in pGV Hm 121 (Ohta et al. 1990) and used to transform *Arabidopsis* C24. T2 homozygous seed of 8-10 independent lines were generated for each construct and analysed for GUS expression using histochemical and quantitative assays as described in the Materials and methods. All four Rbcs-GUS constructs were functional in *Arabidopsis* with expression confined to the green photosynthetic tissues, although the intensity of GUS staining of the leaves was reduced as promoter length decreased, particularly in the shortest of the constructs, PR1. Figure 3A shows the expression of the GUS gene in *Arabidopsis* seedlings for a typical full-length RbcS promoter-GUS line compared to a similar CaMV 35S promoter construct. No obvious expression was observed in the roots of any of the RbcS-GUS *Arabidopsis* plants.

The average GUS enzyme activity in the six most highly expressing independent *Arabidopsis* T2 lines containing each of the promoter constructs is shown in Figure 2C. The highest GUS activity was seen in the PR4 lines which had about twice the average activity of the best six CaMV 35S-GUS lines indicating that it was a strong promoter. The GUS activity in the shorter PR3 promoter-GUS lines was about the same as that of the CaMV 35S lines and approximately 58% of that detected in PR4 promoter plants. Expression levels were even lower with the two shortest constructs. Thus, although we did not study the importance of the different functional motifs identified in the RbcS promoter in any detail, deletion of some of them, particularly the upstream repeated Motif #3, seemed to have significant impacts on the overall activity of the promoter. The shortest promoter examined here retained the key G-box-I-box combinations (Motifs 4+5+2) and, although reduced in overall activity, still retained the green tissue specificity of the full-length promoter.

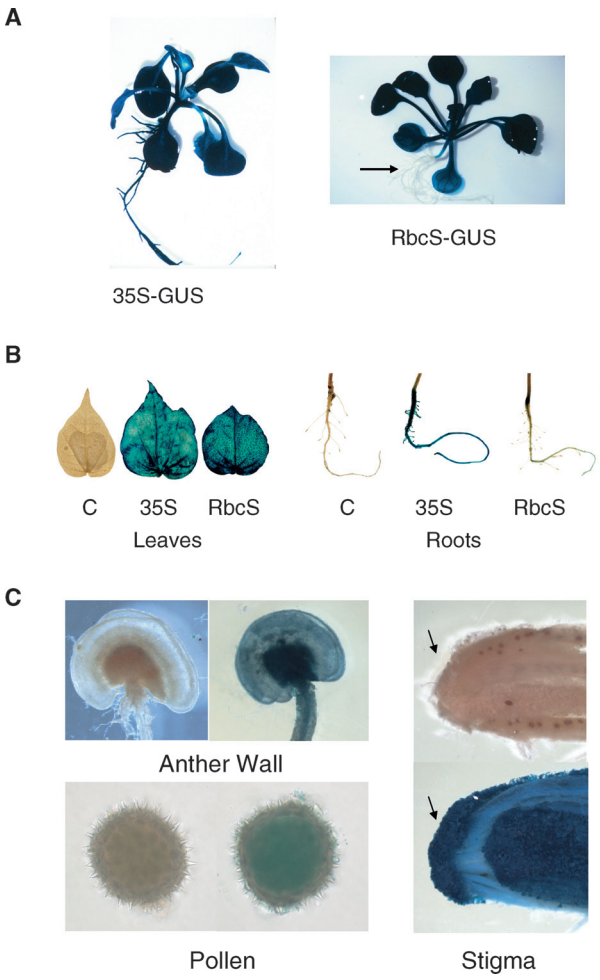

Figure 3. Expression of RbcS-GUS in transgenic cotton and *Arabidopsis*. (A) Typical histochemical staining of transgenic *Arabidopsis* seedlings containing the CaMV 35S-GUS and full-length RbcS-GUS constructs. The arrow indicates the lack of staining in the roots of RbcS-GUS plants. (B) GUS histochemical staining of young homozygous T_2 cotton seedlings containing either the RbcS-GUS or CaMV 35S-GUS constructs and a control non-transformed Coker 315 seedling in leaves (left panel) and roots (right panel). Note the presence of faint blue staining in vasculature of the roots of the RbcS plants. (C) Gus histochemical staining in the anther wall, pollen and stigma of RbcS-GUS plants (right or bottom panels) compared to the non-transformed Coker plant (Left or top panels). Arrows indicate the stigmatic papillae end.

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

Because we had observed the highest activity from the full-length promoter (PR4) and transforming cotton is a difficult and lengthy procedure, only this long construct was introduced into cotton. Twenty fertile primary transformants representing fifteen different

transformation events were generated and all expressed GUS in leaf tissues when stained with X-Gluc. Southern blot analysis of T_0 plants revealed successful integration of the GUS gene into the cotton genome with the number of transgene insertions varying from 1–8 (not shown). Four lines had single copy insertions, 2 lines had two insertions and 14 lines had 3 or more insertions. Plants were taken through to the T_2 generation to assess segregation and to select homozygous lines for more detailed expression analysis and field assessment. Segregation ratios predicted single locus insertions in six independent lines while the rest of the lines were indicative of multiple locus insertions (not shown). Southern blot analysis of DNA isolated from the twelve GUS expressing plants from each line confirmed the stable inheritance of GUS gene in the T_1 generation.

Expression of the RbcS promoter in cotton

To determine the specificity of expression of the RbcS promoter in cotton, we stained different plant parts harvested from a number of T_2 lines as described in the Materials and methods. As indicated by the bright blue precipitate, cotyledons (not shown) and leaves (Figure 3B, left panel) had very high expression of the GUS gene. The roots had almost no detectable GUS expression (although there was a thin band of staining around the vascular bundle, particularly in the highest expressing plants) (Figure 3B, right panel). Young stem and the young floral buds also had low levels of GUS expression (not shown). In mature open flowers staining was observed in the green bract surrounding the flower (not shown), in the green ovary wall (not shown), and in the stigma, and pollen (Figure 3C), but not in ovules (not shown). In some of the highest expressing lines staining was seen in the vasculature of the petal and the filament and at the point of attachment of the anther to the filament and the anther wall (Figure 3C). Plants transformed with a constitutive 35SGUS construct stained strongly in all tissues examined except for the filament and anther wall that were only weakly stained (not shown), while non-transformed control plants showed no staining in any tissue examined (e.g., Figure 3B and 3C). These results are consistent with the data from *Arabidopsis* (Gittins, et al. 2000; Kirby and Kavanagh 2002) and previous studies in cotton (Song et al. 2000) that show that expression of the RbcS promoter is restricted primarily to photosynthetic tissues and especially in leaves. Northern blot analysis of GUS transcripts in different tissues (Figure 4A) confirmed this result. In the leaves of young vegetative plants the highest expressing RbcS-GUS transformant had a steady-state level of GUS mRNA comparable to the highest 35S-GUS line (not shown).

Analysis of the RbcS-GUS plants under Field Conditions

As we were primarily interested in the performance of the RbcS promoter under field conditions, we initiated a small-scale field trial to compare the RbcS-GUS and 35S-GUS plants at a transcriptional level. Ten T_2 plants from each of six single-locus lines of RbcS-GUS (4 single-copy insertions and 2 two-copy insertions, as well as two single-copy 35S-GUS lines and a non-transformed Coker 315 control), were planted in small unreplicated rows. All plants were screened for GUS expression using X-Gluc at the cotyledon stage to verify

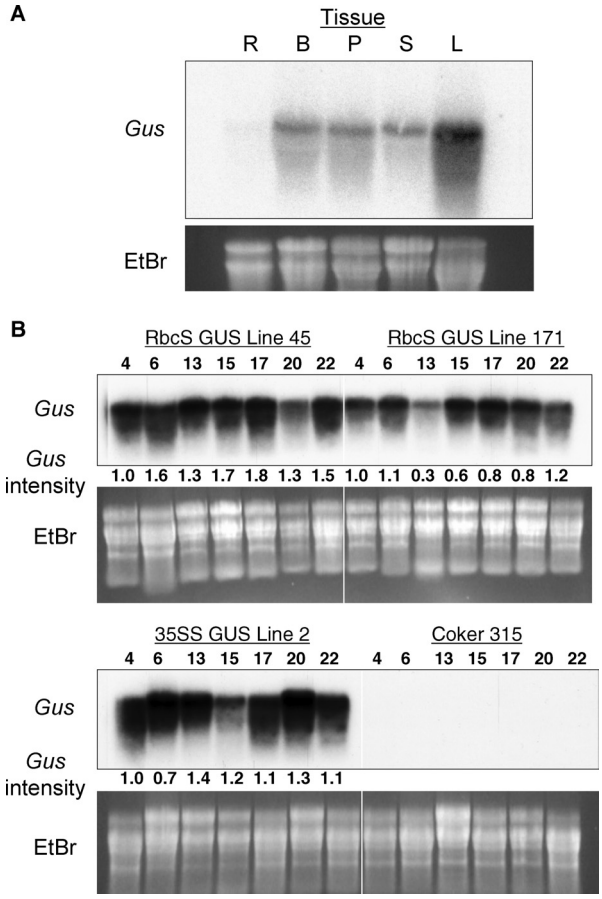

Figure 4. Expression of GUS mRNAs in glasshouse- and field-grown transgenic cotton plants containing the RbcS-GUS construct. (A) Northern blot analysis (upper panel) of GUS mRNAs in different tissues of a mature plant of RbcS-GUS homozygous T_2 cotton line 45 probed with a full-length antisense GUS riboprobe. The lower panel is the ethidium bromide stained gel prior to blotting. Highest expression is in leaves (L) and there is little or no expression in roots (R). Lower levels of expression were seen in green Boll coats (B), Petioles (P) and Stems (S). (B) GUS mRNA levels in field-grown cotton plants containing either the full-length RbcS GUS construct (lines 45 and 171), the CaMV 35S GUS construct (line 2) or a non-transformed Coker 315 control determined by northern blot analysis of samples collected at different times throughout the season. In each panel, sampling times were from left to right, 4, 6, 13, 15, 17, 20 and 22 weeks after planting. The ethidium bromide-stained gel is shown below each blot. Numbers above each lane represent the GUS expression level (corrected for loading) of each sample relative to the 4 week sample for that line averaged from two separate hybridisations.

their zygosity and were sampled periodically up to harvest for analysis of GUS mRNA levels by northern blot analysis. There was no significant drop in mRNA levels for GUS in either the RbcS or the 35S GUS plants (Figure 4B) during the season.

Discussion

In this study, we have used an EST and microarray-based approach to determine the level of abundance, and identify the temporal expression patterns of about 1600 cotton genes from field-grown cotton plants. Since our objective was to identify candidates for promoter isolation, it was important to study the expression patterns of the genes under normal conditions in the field rather than in a temperature-controlled glasshouse. Although plants can be exposed to a number of variables in the field, by careful sampling, pooling and replication, we have been able to obtain consistent results in our microarray experiments. This study thus demonstrates the scope for extending the microarray technology from controlled-growth conditions to field situations to tackle important agronomic questions in cotton.

Based on our analysis of EST abundance, microarray hybridisation signals and northern blot hybridisation analysis, we were able to identify a number of candidate genes from which to isolate strong promoters that could possibly drive expression of transgenes throughout the life cycle of cotton. Although it was perhaps not surprising that these candidates were the highly abundant photosynthetic and housekeeping genes or enzymes of primary metabolism, it was not intuitive that such gene would be stably expressed under field conditions. The most abundant ESTs in the late-season cotton leaf library listed in Table 1 are similar to the list of the 20 most abundant ESTs noted by Nanjo et al. (2004) in stress-treated poplar leaves (dehydration, chilling, salt, heat, ABA and peroxide-treated leaves) probably reflecting the naturally-stressed environment of field-grown irrigated cotton plants that are subject to daily heat and water stresses and natural infestations by pests and diseases.

This study highlighted the high expression and relative stability of the Rubisco small subunit genes as excellent candidates for promoters to drive strong and consistent expression of transgenes in photosynthetic tissues of field grown cotton, provided that a suitable member of this small multigene family was selected. We isolated the promoter of the most highly expressed EST and found it to be identical to a promoter and gene in the Genbank database, and were able to characterise a longer region of that promoter. The promoter carried all the hallmarks of a light-regulated photosynthetic gene and shared a number of motifs with RbcS genes from other plants, as would be expected. When linked to a reporter gene the promoter was able to confer tissue-specific expression and GUS staining was confined to the green photosynthetic tissues in both *Arabidopsis* and cotton. Low expression seen in pollen and roots has been reported for other RbcS promoters (Gittins, et al. 2000; Kirby and Kavanagh 2002) and other photosynthetic gene promoters such as plastocyanin and subunit delta of chloroplast ATP synthase (Bichler & Herrmann 1990), although we cannot discount that some of this expression may be due to residual effects of enhancer elements present in the 35S promoters also on the introduced T-DNA. A shorter promoter region of the cotton RbcS (560 bp) has previously been examined in transgenic cotton (Song et al. 2000), but this analysis was restricted to a couple of T0 plants grown in the glasshouse. In both *Arabidopsis* and cotton our longer promoter construct performed as well as or better than the CaMV 35S promoter and this was maintained over at least the three generations examined, indicating that it would be useful for many applications where expression in the leaves was required, e.g., expressing an insecticidal toxin against lepidopteran insects that feed on green tissues such as leaves, bracts and bolls, or genes for resistance to foliar herbicides. There are few root-chewing pests of cotton so low, or little, expression in the roots is unlikely to be a problem. Under both glasshouse and field conditions the RbcS-GUS construct showed consistent expression in leaves over time and we did not see any significant reduction in transcript abundance post flowering as has been reported for 35S-*Cry1Ac* constructs in cotton (Olsen et al. 2005b). Surprisingly, we also didn't see any significant reduction in the transcripts of GUS from a control CaMV 35S construct grown as a comparator, suggesting that the late season decline noted previously in GM cotton may be a problem specific to expressing Bt toxin genes.

Acknowledgements

We gratefully acknowledge funding for BHRRA and EFN by the Cotton Research and Development Corporation of Australia. We also acknowledge the National Science Foundation and the Plant Genome program for funding a collection of cotton ESTs. The Authors would like to thank Pinghua He, Todd Collins, Kathy Schneebeli, Jane Liu, Merran Hunter and Chris Tyson for their excellent technical assistance and Dr. Iain Wilson for help printing our microarray slides and many helpful suggestions on the microarray experiments. Celia Miller and Carl Davies assisted with photographing the GUS stained tissues.

References

Argüello-Astorga G, Herrrea-Estrella L (1998) Evolution of light-regulated plant promoters. *Ann Rev Plant Physiol Plant Mol Biol* 49: 525–555

Bailey TL, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches.

Bioinformatics 14: 48–54

Bichler J, Herrmann RG (1990) Analysis of the promoters of the single copy genes for plastocyanin and subunit delta of the chloroplast ATP synthase from spinach. *Eur J Biochem* 190: 415–426

Breyne P, De Loose M, Dedonder A, Van Montagu M, Depicker A (1993) Quantitative kinetic analysis of β-glucuronidase activities using a computer-directed microtiter plate reader. *Plant Mol Biol Rep* 11: 21–31

Brown RL, Kazan K, McGrath KC, Maclean DJ, Manners JM (2003) A role for the GCC-box in jasmonate-mediated activation of the PDF1.2 gene of *Arabidopsis*. *Plant Physiol* 132: 1020–1032

Clough SJ, Bent AF (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16: 735–743

Dey N, Maiti IB (1999) Structure and promoter/leader deletion analysis of mirabilis mosaic virus (MMV) full-length transcript promoter in transgenic plants. *Plant Mol Biol* 40: 771–782

Donald RG, Cashmore AR (1990) Mutation of either G box or I box sequences profoundly affects expression from the *Arabidopsis* rbcS-1A promoter. *EMBO J* 9: 1717–1726

Fagard M, Vaucheret H (2000) (Trans)gene silencing in plants: how many mechanisms? *Ann Rev Plant Physiol Plant Mol Biol* 51: 167–194

Fitt GP (2004) Implementation and impact of transgenic Bt cottons in Australia. In: *Cotton Production for the New Millennium*. Proceedings of the third World Cotton Research Conference, Cape Town, South Africa, 9–13 March, 2003, 1778 pages. Agricultural Research Council—Institute for Industrial Crops, Pretoria, South Africa. pp 371–381

Gittins JR, Pellny TK, Hiles ER, Rosa C, Biricolti S, James DJ (2000) Transgene expression driven by heterologous ribulose-5-bisphosphate carboxylase/oxygenase small subunit gene promoters in the vegetative tissues of apple (*Malus pumila mill*.). *Planta* 210: 232–240

Greenplate JT, Mullins JW, Penn SR, Dahm A, Reich BJ, Osborn JA, Rahn PR, Ruschke L, Shappley ZW (2003) Partial characterisation of cotton plants expressing two toxin proteins from *Bacillus thuringiensis*: Relative toxin contribution, toxin interaction, and resistance management. *J Appl Entom* 127: 340–347

Hudson ME, Quail PH (2003) Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. *Plant Physiol* 133: 1605–1616

Kirby J, Kavanagh TA (2002) NAN fusions: a synthetic sialidase reporter gene as a sensitive and versatile partner for GUS. *Plant J* 32: 391–400

Lazo GR, Stein PA, Ludwig RA (1991) A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Bio/technology* 9: 963–967

Li Z, Jayasankar S, Gray DJ (2001) Expression of a bifunctional green fluorescent protein (GFP) fusion marker under the control of three constitutive promoters and enhanced derivatives in transgenic grape (*Vitis vinifera*). *Plant Sci* 160: 877–887

Manzara T, Carrasco P, Gruissem W (1991) Developmental and organ-specific changes in promoter DNA-protein interactions in the tomato RbcS gene family. *Plant Cell* 3: 1305–1316

Murray F, Llewellyn D, McFadden H, Last D, Dennis ES, Peacock WJ (1999) Expression of the *Talaromyces flavus* glucose oxidase gene in cotton and tobacco reduces fungal infection, but is also phytotoxic. *Mol Breeding* 5: 219–232

Nanjo T, Futamura N, Nishiguchi M, Igasaki T, Shinozaki K, Shinohara K (2004) Characterization of Full-length Enriched Expressed Sequence Tags of Stress-treated Poplar Leaves. *Plant Cell Physiol* 45: 1738–1748.

Odell JT, Nagy F, Chua N-H (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* 313: 810–812

Ohta S, Mita S, Hattori T, Nakamura K (1990) Construction and expression in tobacco of a beta glucuronidase (GUS) reporter gene containing an intron within the coding sequence. *Plant Cell Physiol* 31: 805–813

Olsen KM, Daly JC, Finnegan EJ, Mahon RJ (2005a) Changes in *Cry1Ac* Bt Transgenic Cotton in Response to Two Environmental Factors: Temperature and Insect Damage. *J Econ Entomol* 98: 1382–1390

Olsen KM, Daly JC, Holt HE, Finnegan EJ (2005b) Season-long Variation in Expression of the *Cry1Ac* Gene and Efficacy of Bt Toxin in Transgenic Cotton against *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae). *J Econ Entomol* 98: 1007–1017

Potenza C, Aleman L, Sengupta-Gopalan C (2003) Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation. *In Vitro Cell Devel Biol—Plant* 40: 1–22

Rungis D, Llewellyn D, Dennis ES, Lyon BR (2005) Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars of Australian and American origin. *Austr J Agric Res* 56: 301–307

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual.*, Cold Spring Harbor, Cold Spring Harbor Laboratory Press

Sanger M, Daubert S, Goodman RM (1990) Characteristics of a strong promoter from the figwort mosaic virus: comparison with the analogous 35S promoter from cauliflower mosaic virus and the regulated mannopine synthase promoter. *Plant Mol Biol* 14: 433–443

Schenk PM, Remans T, Sagi L, Elliott AR, Dietzgen RG, Swennen R, Ebert PR, Grof CPL, Manners JM (2001) Promoters for pregenomic RNA of banana streak badnavirus are active for transgene expression in monocot and dicot plants. *Plant Mol Biol* 47: 399–412

Schunmann PHD, Llewellyn DJ, Surin B, Boevink P, De Feyter RC, Waterhouse, PM (2003a) A suite of novel promoters and terminators for plant biotechnology. *Funct Plant Biol* 30: 443–452

Schunmann PHD, Surin B, Waterhouse PM (2003b) A suite of novel promoters and terminators for plant biotechnology—II. The pPLEX series for use in monocots. *Funct Plant Biol* 30: 453–460

Small RL, Ryburn JA, Wendel JF (1999) Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). *Mol Biol Evol* 16: 491–501

Song P, Heinen JL, Burns TH, Allen RD (2000) Expression of two tissue-specific promoters in transgenic cotton plants. *J Cotton Sci* 4: 217–223

Sunilkumar G, Mohr L, Lopata-Finch E, Emani C, Rathore KS (2002) Developmental and tissue-specific expression of CaMV 35S promoter in cotton as revealed by GFP. *Plant Mol Biol* 50: 463474

Townsend BJ, Llewellyn DJ (2002) Spatial and temporal regulation

of a soybean (*Glycine max*) lectin promoter in transgenic cotton (*Gossypium hirsutum*). *Funct Plant Biol* 29: 835–843

Udall JA, Swanson J, Haller K, Rapp RA, Sparks ME, Hatfield J, Yu Y, Wu Y, Arpat AB, Sickler BA, Wilkins TA, Guo J-Y, Chen X-Y, Scheffler J, Taliercio E, Turley R, McFadden H, Payton P, Allen R, Zhang D, Haigler C, Wilkerson C, Suo J, Schulze SR, Pierce ML, Essenberg M, Kim H, Llewellyn DJ, Dennis ES, Kudrna D, Wing R, Paterson AH, Soderlund C, Wendel JF (2006) A global assembly of cotton ESTs. *Genome Res* 16: 441–450

Verdaguer B, de Kochko A, Beachy RN, Fauquet C (1996) Isolation and expression in transgenic tobacco and rice plants, of the cassava vein mosaic virus (CsVMV) promoter: *Plant Mol Biol* 31: 1129–1139

Wan CH, Wilkins TA (1994) A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Anal Biochem* 223: 7–12

Wilson DL, Buckley MJ, Helliwell CA, Wilson IW (2003) New normalization methods for cDNA microarray data. *Bioinformatics* 19: 1325–1332

Wilson IW, Kennedy GC, Peacock WJ, Dennis ES (2005) Microarray analysis reveals vegetative molecular phenotypes of *Arabidopsis* flowering time mutants. *Plant Cell Physiol* 46: 1190–1201

Wu Y, Rozenfeld S, Defferrard A, Ruggiero K, Udall JA, Kim H-R, Llewellyn, DJ, Dennis ES (2005) Cycloheximide treatment of cotton ovules alters the abundance of specific classes of mRNAs and provides a method of generating novel ESTs for microarray expression profiling. *Mol Genet Genomics* 274: 477–493

Wu Y, Machado A, White RG, Llewellyn DJ, Dennis ES (2006) Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *Plant Cell Physiol* 47: 107–127