Theor Appl Genet (1999) 98: 757–764

S. J. Orford · T. J. Carney
N. S. Olesnicky · J. N. Timmis

Characterisation of a cotton gene expressed late in fibre cell elongation

Received: 18 August 1998 / Accepted 18 September 1998

Abstract Full-length transcripts corresponding to a cotton fibre-specific cDNA, pFS18, were isolated by cDNA library screening and 5′-RACE. The cDNA sequences contained a short open reading frame of only 71 amino acids which showed a limited amount of similarity to lipid transfer proteins. Southern analysis revealed the presence of 4–6 related genes in the tetraploid (AD) *G. hirsutum* genome, one of which, *FS18A*, was isolated and shown to contain a short open reading frame which originated from the ancestral A diploid genome. Sequence differences between the gene and the full-length pFS18 cDNA, together with the finding that pFS18 is transcribed from the D diploid genome, demonstrated that *FS18A* does not encode the abundant fibre-specific transcripts represented by pFS18. However, RT-PCR experiments showed that both pFS18 and *FS18A*-derived transcripts are present in fibre cells at elevated levels compared with other cotton tissues, with the accumulation of transcripts from both genes being greatest during and after the onset of secondary wall synthesis.

Key words *Gossypium hirsutum* · cDNA sequence · Differential gene expression · Cotton fibre

Introduction

The cotton fibre of agricultural significance is the product of cultivated *Gossypium* species, of the Malvaceae family. Fibres arise from single epidermal cells of cotton ovules and begin to differentiate at about the time of anthesis. In addition to their economic importance, cotton fibres are an ideal experimental system in which to study cellular and developmental events, as cotton fibre development is characterised by a precise and synchronous growth pattern which is uncomplicated by cell division. Since the secondary cell wall of cotton fibres consists almost entirely of cellulose, they also provide an excellent system in which to study cellulose biosynthesis.

Numerous investigations at both the ultrastructural and biochemical level have shown that fibre cells develop in four distinct but overlapping phases. An initiation period is followed by extensive cell elongation (primary wall synthesis) and secondary wall deposition, which begins 16–19 days post-anthesis (DPA) and continues until the cotton fibre is mature (Basra and Malik 1984). The control of fibre development and cellulose biosynthesis is complex and not entirely understood, being influenced by both genotypic and environmental factors. With increasing demand for high quality, natural fabrics, improvements to fibre properties and yield would greatly benefit the cotton industry worldwide. Improvement of cotton fibres using molecular genetic techniques is feasible (John and Stewart 1992; John and Keller 1996) but requires an understanding of fibre developmental processes and a characterisation of the genes and regulatory proteins involved.

The number of genes active in the developing cotton fibre is similar to that in other plant organs, and consists of both general and fibre-specific genes (Graves and Stewart 1988; John and Stewart 1992). Working on the assumption that genes expressed specifically in cotton fibre cells will be important in their development, several groups have recently reported the isolation of fibre-specific genes and cDNAs. Some of the genes encode homologues of previously characterised proteins, such as lipid transfer proteins (LTPs)

Communicated by R. Hagemann

S. J. Orford (✉) · T. J. Carney · N. S. Olesnicky · J. N. Timmis
Department of Genetics, The University of Adelaide,
Adelaide 5005, Australia
Tel.: + 61-8-303 3013
Fax: + 61-8-303 4399
E-mail: sorford@genetics.adelaide.edu.au

(Ma et al. 1995, 1997; Orford and Timmis 1997), tubulin isoforms (Dixon et al. 1994), a GTPase (Delmer et al. 1995) and an acyl carrier protein (Song and Allen 1997). The majority, however, showed no similarity to known sequences and encode novel proteins of undetermined function (John and Crow 1992; John 1995; John and Keller 1995; Rinehart et al. 1996).

Previously, we described the isolation and partial characterisation of fibre cDNAs which, as identified by a differential screen, were specific to the developing cotton fibre (Orford and Timmis 1997). One such clone, pFS18, was only 98 bp in length and screening of nucleotide and protein databases did not reveal any significant homologies to known sequences. Northern-blot analysis showed that pFS18 was unique among the fibre-specific clones in that transcript accumulation was greatest in fibres aged 10–18 DPA, implying a role for the encoded protein in the initiation of secondary wall deposition (Graves and Stewart 1988; Orford and Timmis 1997). We now describe the isolation and characterisation of full-length pFS18 mRNAs and a gene homologue, *FS18A*, from the genome of *Gossypium hirsutum*, cv Siokra 1-2.

Materials and methods

Plant material and growth conditions

Seeds of *G. hirsutum* L. cultivars Siokra 1-2, Siokra 1-4, Siokra L-22 and CS7S [$(AD)_1$ genome] were obtained from Cotton Seed Distributors, Wee Waa, NSW. Seeds of *Gossypium barbadense* L. cv Pima S-6 ($[AD]_2$ genome) were provided by R. McAllum ("Tandou", Menindee, NSW). Leaves of *Gossypium herbaceum* L. (A_1 genome) and genomic DNA preps of *Gossypium raimondii* Ulbr. (D_5 genome) were obtained from Dr. C. Brubaker (CSIRO Division of Plant Industry, Canberra, ACT). Cotton seeds were sown in soil and grown in growth cabinet conditions (day/night temperatures of 30°C/25°C, light/dark cycle of 16/8 h). Flowers were tagged on the day of anthesis, and fibres of appropriate ages were manually harvested and frozen in liquid nitrogen.

DNA and RNA isolation

Cotton genomic DNA was isolated from leaf tissue by the method of Dellaporta et al. (1985). Total RNA was isolated from tissues of *G. hirsutum* L. cv Siokra 1-4 (Wan and Wilkins 1994) and stored in aliquots under ethanol at −80°C . Plasmid DNA was routinely purified using the alkaline-lysis method (Sambrook et al. 1989).

Construction and screening of cotton DNA libraries

A 13-DPA fibre cDNA library consisting of 1.1×10^4 clones was constructed, plated and screened with the pFS18 cDNA as previously described (Orford and Timmis 1997).

To construct a genomic library, cotton DNA (cv Siokra 1-2) was partially digested with *Sau*3AI and size-fractionated by sucrose-gradient centrifugation (Ausubel et al. 1989). Sucrose-gradient fractions containing DNA fragments in the size range 15–23 kb (mean size 19 kb) were pooled and the DNA was ligated (Sambrook et al. 1989) to *Bam*HI-digested, de-phosphorylated λGEMII arms (Promega). The recombinant phages were packaged using Packagene (Promega) and plated at high density with *Escherichia coli* KW251 cells. The genomic library, consisting of 1×10^5 clones, was lifted in duplicate onto Hybond-N^+ membranes (Amersham) and screened with a 616 bp pFS18 cDNA. DNA from positively hybridising plaques was purified using the plate lysate method of Sambrook et al. (1989) and genomic DNA fragments subcloned (Sambrook et al. 1989) into pBluescript SK(−) (Stratagene).

Southern blotting and hybridisation

DNA samples were digested with restriction endonucleases under conditions recommended by the supplier (Boehringer). Following electrophoresis, DNA was transferred to a Hybond-N^+ membrane (Amersham), following the manufacturer's instructions.

Plasmid inserts requiring radioactive labelling were purified from an agarose gel and oligolabelled (Hodgson and Fisk 1987) with α-[^{32}P]dATP (Amersham). Pre-hybridisation, hybridisation and washes of Southern blots and plaque lifts were carried out as described previously (Orford and Timmis 1997).

Sequencing of double-stranded DNA templates

Genomic subclones were sequenced using oligonucleotide primers designed to the genomic sequence. Plasmids were sequenced using the dideoxy chain-termination method as previously described (Orford and Timmis 1997). Automated sequencing (ABI Prism Model 377) was performed using dye-terminator cycle sequencing reactions (Applied Biosystems-Perkin Elmer) and sequence analysis utilised the GCG Sequence Analysis Software Package version 8 (Genetics Computer Group, Madison, Wis., 1984).

Rapid amplification of cDNA ends (5′-RACE)

The 5′ end of the pFS18 cDNA was obtained using the 5′-AmpliFINDER™ RACE kit (Clontech Laboratories) as described previously (Orford and Timmis 1997), using two specific antisense primers, 5′-GTG GCA TCA GCA CCA TCA CCA TTC C-3′ (for reverse transcription) and 5′-AAG GAA TTC GCA CAA CAC CAA CAG ACA TGC AAG C-3′ (for PCR).

PCR of genomic LTP sequences

Genomic DNA templates were amplified using primers specific for the pFS18 genes; namely, PFS18-5′B, TGT GCC ATC GTC ATG AAT GC and PFS18-3′; 5′-ATG ATG GTT GGA TAC CCA GG-3′ (for pFS18), and primers PFS18-5′; 5′-TGC GAC ATC GTC AGG GGT CT-3′ and PFS18-3′ (for gene FS18A). Each reaction contained approximately 100 ng of genomic DNA or 20 pg of clone DNA, 0.4 mM of each dNTP, 1 μM of each primer, 2 mM of $MgCl_2$ and 1.5 units of *Taq* DNA Polymerase (Bresatec) in a final volume of 25 μl of 1 × PCR reaction buffer. Initial template denaturation was at 94°C for 4 min followed by 25 PCR cycles of denaturation at 94°C for 30 s, primer annealing at 55°C for 30 s and extension at 72°C for 1 min.

Reverse transcription-PCR

Total RNA (3 μg) was reverse transcribed using poly(T) (Pharmacia) and M-MLV reverse transcriptase under conditions specified by the enzyme supplier (BRL). Following reverse transcription (RT),

a 1/300 sample (for pFS18) or a 1/30 sample (for *FS18A*) of the single-stranded cDNA product was amplified by PCR under conditions, and using the primer combinations, described above for genomic DNA. Actin transcripts were amplified from a 1/300 sample of RT products using the PCR primers of Shimizu et al. (1997), under the conditions outlined above except with an annealing temperature of 61°C. Replicate RT reactions without reverse transcriptase were included to ensure that PCR products were not the result of amplification of contaminating DNA.

Results

Characterisation of the full-length cDNA

Probing of the 13-DPA fibre cDNA library with the 98 bp fibre-specific cDNA clone, pFS18 (Orford and Timmis 1997), identified several positive clones, of which four, pFS18-C, pFS18-G, pFS18-W and pFS18-X, provided sequences of 616 bp, 610 bp, 556 bp and 587 bp, respectively. The sequences were identical to each other in regions of overlap both with respect to each other and to the original 3′ 98 bp of pFS18, but each differed in their length at the 5′ end.

To determine whether a full-length pFS18 transcript had been obtained, rapid amplification of cDNA ends (5′-RACE) was carried out, using oligonucleotide primers within the extended pFS18 cDNA sequence. A major PCR product of approximately 190 bp and a minor product of 270 bp were observed, both of which hybridised to the longest existing cDNA probe (data not shown). These were cloned to generate clones RACE1 and RACE10, corresponding to the major and minor 5′-RACE PCR products, respectively. Both clones contained sequences which overlapped the cDNA sequence and were identical to each other and to the pFS18 clones in the regions of overlap. The 5′-RACE clones were terminated at the 5′ end by a G residue, indicative of the presence of the 7-methylguanosine mRNA cap which can be reverse transcribed during a 5′-RACE extension (Hirzmann et al. 1993). The RACE1 insert, of 122 bp, extended the longest existing sequence by 23 bp whilst RACE10 (146 bp) extended the sequence by 47 bp to result in full-length sequences of 641 bp and 662 bp, respectively, excluding the 3′ poly(A) tail (Fig. 1). This is in good agreement with a transcript size of 700 nt estimated from a Northern blot (Orford and Timmis 1997). The two different clones may reflect alternative transcription start sites, alternative splicing of a single transcript, or be the result of the detection of transcripts from two distinct but related genes.

The 662 bp sequence of the full-length pFS18 cDNA (Fig. 1) contains a single significant open reading frame potentially encoding a polypeptide of 71 amino acids and includes an unusually long 3′ untranslated region (UTR) of 360 bp. Numerous stop codons were present in all three reading frames. After the ATG translation initiation codon, there is a stretch of 26

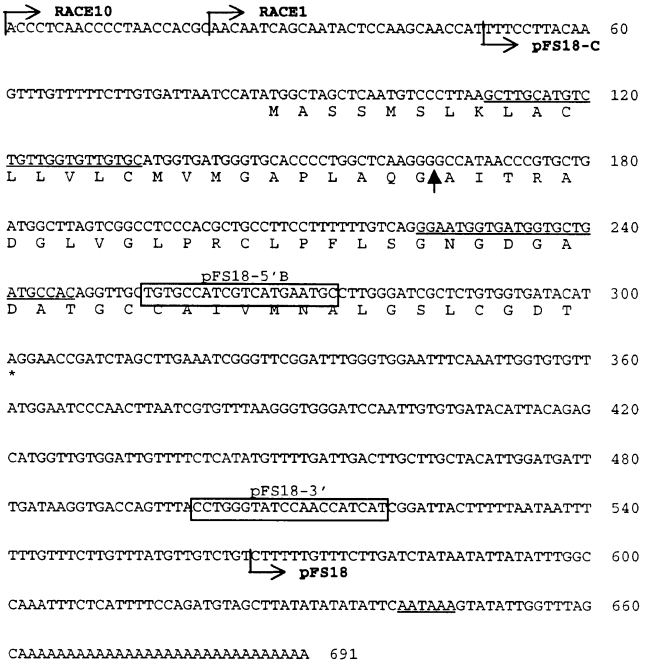

Fig. 1 Nucleotide and deduced amino-acid sequence of the full-length pFS18 cDNA, constructed from pFS18-C and 5′-RACE clones. The predicted amino-acid sequence of the longest ORF is shown in a single letter code below the DNA sequence. The putative signal peptide cleavage site, as predicted according to von Heijne (1983), is indicated by an *arrow*, as are the 5′ boundaries of the pFS18-C and pFS18 sequences. Positions of PCR primers are *boxed* and the regions homologous to the oligonucleotide primers used for 5′-RACE are *underlined*. A putative polyadenylation signal in the 3′-UTR is also *underlined*

amino acids which contains many of the features characteristic of signal peptides (Nothwehr and Gordon 1990) and processing of the conceptual pre-protein (Fig. 1) would generate a mature peptide of only 45 amino acids. A comparison of the full-length pFS18 cDNA sequence with sequences in the databases (NCBI non-redundant, March 1998) revealed significant similarity to lipid transfer proteins from cotton. At the nucleotide level, the pFS18 sequence showed 81% and 80% identity in 229-bp overlaps to the 5′ ends of fibre-specific cDNAs GH3 (Ma et al. 1995) and pFS6 (Orford and Timmis 1997), respectively, which encode LTPs. In addition, the sequence was 76% identical in a 196-bp overlap with an LTP gene, *ltp6* (Ma et al. 1997), but the similarity did not extend beyond the extreme 5′ sequence of any LTP gene.

At the protein level, the pFS18 conceptual translation was similar to LTPs from a variety of plants, with the region of similarity largely confined to the 26 residue transit peptide and sharing 88%, 85% and 73% amino-acid identity with the cotton LTP sequences GH3 (Ma et al. 1995), pFS6 (Orford and Timmis 1997) and *ltp6* (Ma et al. 1997), respectively. A lesser degree of similarity was observed to signal peptides of LTP homologues from spinach (Bernhard et al. 1991),

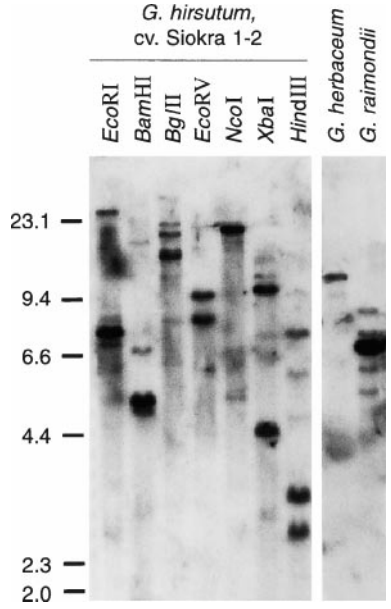

Fig. 2 Cotton genomic DNA analysis for the pFS18 fibre-specific cDNA. Each lane was loaded with 10 µg of total genomic DNA as indicated. *Hin*dIII and *Bam*HI contain recognition sites within the cDNA sequence. *Arrows* on the left indicate molecular-weight markers with sizes in kb. *G. herbaceum* and *G. raimondii* DNA samples were both restricted with *Eco*RI

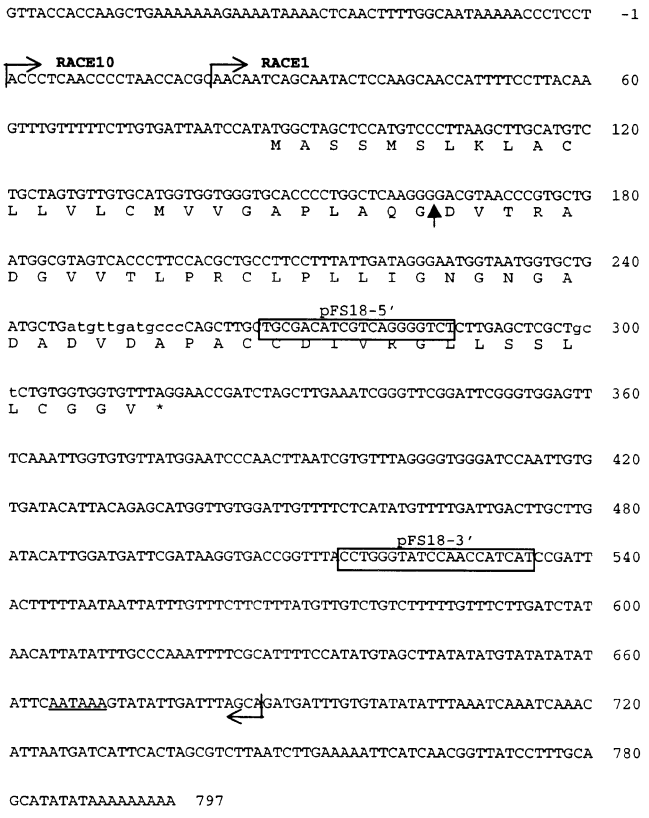

Fig. 3 Nucleotide and deduced amino-acid sequence of the gene *FS18A*. The predicted amino-acid sequence of the longest ORF is shown in *single letter code* below the DNA sequence and insertions in the protein-coding region are in *lower case lettering*. The putative signal peptide cleavage site is indicated by an *arrow*, positions of PCR primers are *boxed* and a putative polyadenylation signal in the 3′-UTR is *underlined*. The sites of transcription initiation and termination, as determined by alignment with the cDNA, are also indicated by *arrows*

almond, tomato (Torres-Schumann et al. 1992) and tobacco (Masuta et al. 1992), with 54%, 50%, 42% and 42% identity, respectively. The similarity extended into the putative mature peptide encoded by pFS18, with the 31 N-terminal residues sharing between 29% and 39% sequence identity to plant LTPs including those from *G. hirsutum*, but the pFS18 protein was missing one of the four highly conserved cysteine residues present in this region of LTPs.

Southern blot analysis

Probing of a Southern blot of genomic DNA from several allotetraploid (AD) cotton cultivars, as well as from two related diploid species, *G. herbaceum* L. (genome designation A_1) and *G. raimondii* Ulbr. (genome designation D_5), with the full-length pFS18 cDNA identified one or two major and a few minor bands of hybridisation per lane (Fig. 2), indicative of a low copy number of pFS18 homologues in the tetraploid cotton genome. *Eco*RI digestion of Siokra 1-2 genomic DNA produced a single prominent hybridising fragment sized approximately 7.7 kb which was also present in DNA from other commercial tetraploid cotton varieties, *G. hirsutum*, cv Siokra 1-4, Siokra L-22 and CS7S and *G. barbadense*, cv Pima S-6, but was larger (8.5 kb) in some varieties (data not shown).

A strongly hybridising fragment and four minor fragments were evident in DNA from *G. raimondii* (Fig. 2), the D genome diploid, only one of which was similar in mobility to its counterpart in *G. hirsutum* (Fig. 2). DNA from the A genome diploid, *G. herbaceum*, showed only weak hybridisation to the pFS18 cDNA probe (Fig. 2) and therefore contains more distantly related sequences.

The sequence of a pFS18-like gene

Screening of a *G. hirsutum* genomic library with the full-length pFS18 cDNA detected 13 positive clones, a number of which had been previously identified as LTP homologues. Restriction analysis of two novel clones, λFS18(O) and λFS18(Q), showed that they were overlapping, with a single common region hybridised by pFS18, designated *FS18A* (data not shown). Sequencing of subclones containing the candidate gene produced 1037 bp of nucleotide sequence, the relevant region of which is shown in Fig. 3. Numerous stop codons were present in all six open reading frames, the

longest open reading frame being of 76 amino acids, the sequence of which could be aligned with that of the cDNA.

The sequence of *FS18A* showed 94% nucleotide identity to the pFS18 cDNA in the region of overlap, with 89% identity in the probable protein-coding region. Compared with the cDNA there were two in-frame insertions, one of 12 bp and one of 3 bp, in *FS18A* which increased the length of the conceptual translation product by five amino acids (Fig. 3). Comparison of the deduced amino-acid sequence of the protein encoded by *FS18A* with that of the cDNA revealed 76% identity, not including the two in-frame insertions, with only 8 of the 17 differences involving amino acids with similar chemical properties. As for the cDNA sequence, the gene sequence showed significant similarity to the three published cotton LTPs over the first 200 bp of the transcribed region of the gene, showing 84% identity (206 bp overlap), 81% identity (213 bp overlap) and 69% identity (285 bp overlap) with pFS6, GH3 and *ltp6*, respectively, but there was little or no further similarity.

Genomic origins of the pFS18-like genes

The presence of a strongly-hybridising fragment in the genome of *G. raimondii* and the absence of such a fragment from *G. herbaceum* (Fig. 2) suggested that the pFS18 gene is derived from the ancestral diploid D genome. To test this hypothesis, PCR was performed on genomic DNA with primers specific for the pFS18 cDNA sequence. Primer design utilised the 12-nucleotide insertion in *FS18A* (Fig. 3) which is absent from pFS18 (Fig. 1). An amplification reaction using primers PFS18-5′B and PFS18-3′ yielded a fragment of 265 bp, the expected size for the gene which encodes pFS18, from the genomes of the A-D allotetraploids *G. hirsutum* cv Siokra 1-2 and Siokra 1-4 and *G. barbadense* cv Pima S-6 and the D genome diploid *G. raimondii* (Fig. 4 A). A corresponding amplimer was absent when DNA from the A genome diploid *G. herbaceum* was used as a template, implying that the gene encoding the fibre-specific cDNA, pFS18, in tetraploid cotton is derived from the ancestral D genome, a result consistent with the Southern analysis.

Amplification reactions using a 5′ primer specific to *FS18A* sequences yielded a fragment of 268 bp from the genomes of the A-D allotetraploids (Fig. 4 A). However, in contrast to the result with pFS18-directed primers, a corresponding fragment was amplified from DNA of the A genome diploid *G. herbaceum*, but was absent for the D genome diploid *G. raimondii* DNA (Fig. 4 A). This result implies that gene *FS18A* from tetraploid cotton is derived from the ancestral A genome. Amplification reactions using λFS18(Q) and pFS18-G cloned DNAs as templates showed that the primers were specific for each pFS18-related gene

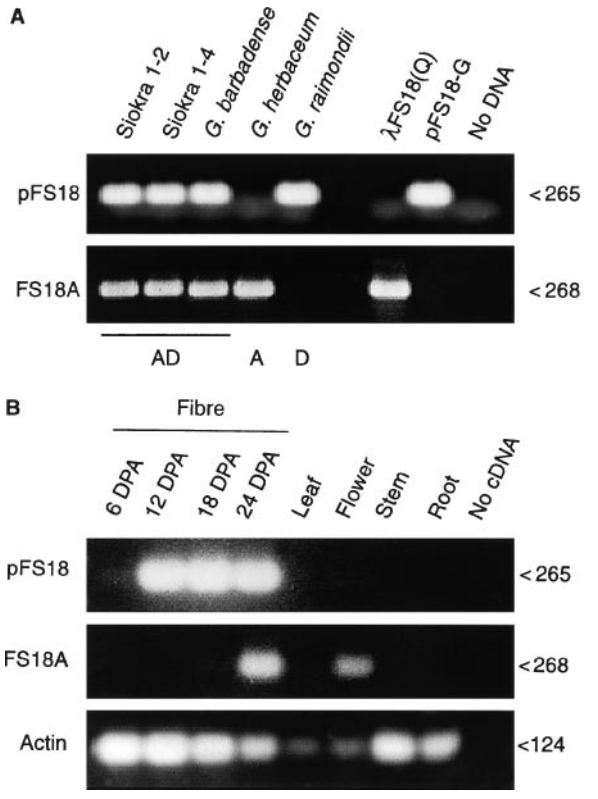

Fig. 4 **A** PCR analysis of two pFS18-like genes in *G. hirsutum*. Each lane contains equivalent loadings of PCR products obtained from DNA templates as indicated and electrophoresed in 1.5%(w/v) agarose. The *numbers* to the right of the figure indicate the size of the PCR products in bp and the genome constituency of each species is denoted below the gel. **B** RT-PCR analysis of expression patterns of cotton pFS18-like genes. Each lane contains equivalent loadings of PCR products obtained from the reverse transcription of cotton tissue RNA and electrophoresed in 1.5%(w/v) agarose. The template concentration (relative to that for pFS18) was 10 for both *FS18A* and actin. The *numbers* to the right of the figure indicate the size of the PCR products in bp

(Fig. 4 A). This result, combined with the observed sequence differences between the pFS18 cDNA and *FS18A*, indicates that the latter gene does not encode the abundant fibre-specific mRNA isolated as pFS18 (Orford and Timmis 1997).

Expression of two pFS18-like genes in cotton

In order to examine the expression patterns of the two distinct pFS18-like genes, an RT-PCR experiment was carried out, which utilised RNA templates from fibre ages 6, 12, 18 and 24 DPA as well as RNAs from leaf, flower, stem and root tissue. As shown in Fig. 4 B, pFS18-like transcripts are more readily amplified by both primer pairs from developing cotton fibres, with only low levels of transcripts detectable in the remaining tissues (Fig. 4 B). A small amount of amplification was observed from whole-flower tissue for both primer sets.

Accumulation of pFS18 transcripts in fibres appeared to be greatest immediately prior to the onset of secondary wall synthesis, at 12–18 DPA, a result consistent with that obtained from Northern analysis (Orford and Timmis 1997). Expression of the *FS18A* gene appears to be distinct from that of pFS18, with higher transcript accumulation after the onset of secondary wall synthesis (16–18 DPA) than during the stage of elongation (0–18 DPA). In addition, the expression of the gene encoding pFS18 transcripts appears to be higher than that of *FS18A*, as 10-fold more reverse transcribed template was required for detection of a product in the latter reaction. The presence of more template in the *FS18A* reactions also explains the seemingly higher levels of flower transcripts (Fig. 4 B).

Discussion

The pFS18 fibre-specific cDNA clone was used to isolate additional cDNA clones of which four were analysed in detail, with the resultant sequence used as a basis for 5′-RACE PCR. The full-length cDNA appears to contain only one significant open reading frame, encoding a putative polypeptide of 71 amino acids. The unusual structure of the pFS18 mRNA, namely a short protein and a long 3′ UTR, was evident in each of four independently isolated cDNAs and in a closely related gene, *FS18A*, isolated from the cotton genome.

Considerable differences were observed between the nucleotide sequences and conceptual translations of the cDNA and the related gene, with only 76% amino-acid identity and two in-frame insertions found in the gene. *G. hirsutum* is a 1–2 million year-old allotetraploid which probably arose by hybridisation between diploid donors of its constituent A and D genomes, followed by a polyploidisation event. The extant diploids *G. herbaceum* and *G. raimondii* provide candidate models for the A and D subgenomes, respectively (Endrizzi et al. 1985), and these species were used as representatives of the ancestral genomes. A PCR experiment showed that the two genes originate from different diploid genomes, with *FS18A* being derived from the ancestral A genome and the pFS18 gene from the ancestral D genome. The sequence differences and the different diploid genomic origins indicate that *FS18A* does not encode the fibre-specific pFS18 cDNA isolated previously (Orford and Timmis 1997) but that it is a close homologue. Therefore, there are at least two genes present in tetraploid cotton, one of which encodes the abundant cDNA, pFS18, and the gene *FS18A* which we describe herein. The sequence of the *FS18A* gene does not help to distinguish between the two possible origins for the isolation of two 5′-RACE PCR products, as it could encode both mRNA species (Figs. 1, 3).

RT-PCR experiments showed that both genes are expressed in cotton fibres at elevated levels compared with other cotton tissues and that each is under developmental control, with very distinct patterns of temporal expression during fibre growth. *FS18A* appears to be expressed late in fibre elongation, during and after the time when secondary wall synthesis has commenced, whilst pFS18-derived transcripts were also detected prior to the onset of secondary wall synthesis, at 12–18 DPA. By 16–18 DPA, the rate of fibre elongation is decreasing and proteins important in secondary wall deposition are increasing in expression (Graves and Stewart 1988). The pattern of transcript accumulation for pFS18 is similar to that observed for a fibre-enhanced GTPase (Delmer et al. 1995), an enzyme thought to have a role in the cytoskeletal reorganisation which occurs during the transition between primary and secondary wall synthesis, and it is possible that the pFS18-encoded protein also has a role during this transition period. The FS18A gene is activated during a critical period in fibre development, when extensive amounts of cellulose are synthesised and deposited in the secondary cell wall, and the encoded protein is therefore unlikely to be involved in primary cell wall synthesis or fibre elongation. The presence of the signal peptide indicates that the proteins are secreted and implies a structural or enzymatic role for the small pFS18-like mature proteins in the cell wall or the extracellular matrix.

It is unknown whether the pFS18 and *FS18A* transcripts are translated *in vivo*. The position of the functional initiating codon is assumed to be the first AUG in each case, since nucleotides in the vicinity of the start codon best resemble the consensus translation start site 5′-TAA ACA *ATG* GCT-3′ and the "first AUG" rule holds for the vast majority of plant genes (Joshi 1987). Additional candidates for initiation codons are in a sub-optimal sequence context and are located downstream from the putative functional AUG (Figs. 1, 3). The conceptual translations of both pFS18 and *FS18A* sequences are terminated by a TAG stop codon, which is an uncommon stop codon in plant genes (Murray et al. 1989), including *G. hirsutum* (Nakamura et al. 1996). A second TAG codon is present 9 bp downstream from the first (Figs. 1, 3) and both TAG termination codons are in an optimal sequence context, being followed by a G residue (Brown et al. 1990), which suggests that alternative translation or frameshifting is unlikely to generate a longer polypeptide. The preferred stop codon for plant genes, TAA, is present in-frame a further 60 bp downstream, implying that even if a certain amount of readthrough occurred, the mature polypeptides encoded by pFS18 and *FS18A* would still only be 69 and 74 residues in length, respectively.

The nucleotide sequences of the pFS18 cDNA and the FS18A gene have a high level of sequence similarity

(up to 85%) at their 5′ ends, particularly in the 5′ UTR and signal sequence, to LTPs from a number of plants. Lipid transfer proteins are small proteins of approximately 120 amino acids, containing 6–8 conserved cysteine residues which are thought to be necessary for tertiary structure formation (for a review see Kader 1996). LTPs have been characterised in a number of plants and are encoded by multigene families, the members of which often display distinct and tissue-specific expression patterns. All LTPs characterised to date are synthesised as precursors with N-terminal signal peptides and secretion of the protein has been demonstrated in several cases, with localisation to the cell wall (Bernhard et al. 1991; Sterk et al. 1991; Thoma et al. 1993; Pyee et al. 1994). In cotton, the three fibre-specific LTPs characterised to date are thought to have a role in the deposition of cutin monomers within the cuticular layer of the cell wall.

However, considering the highly conserved nature of LTPs in other plants, it is very unlikely that the pFS18 cDNA clone encodes an LTP. The conceptual translation products of pFS18 and *FS18A* do not contain the LTP signature, including only three of eight highly conserved cysteine residues, and, at 71 and 76 amino acids, are considerably shorter than any known LTP. The two gene types utilise a strikingly similar transit peptide but the encoded proteins are exported from the fibre cells at very different times in fibre development. Expression of the cotton LTP genes peaks during the elongation phase of fibre development, at a time when the cuticle is actively deposited, whereas transcript accumulation of both pFS18-like genes peaked later, one at 12–18 DPA and the other at 18–24 DPA (Fig. 4).

Characterisation of the genes involved in fibre development and the concomitant isolation of their regulatory sequences represent an important advance towards the genetic manipulation of cotton fibre development and an improvement of fibre quality. This paper describes the observation of a fibre-specific cDNA which, based on its lack of similarity to known sequences, encodes a novel, short polypeptide of unknown function. Significantly, the expression pattern of the pFS18 genes in fibre development is unique among the fibre-specific genes thus far isolated in our laboratory, since all others are expressed maximally during fibre cell elongation or primary cell wall synthesis (Orford and Timmis 1997, 1998). The isolation of the *FS18A* promoter could provide a powerful tool in the expression of heterologous genes specifically during the secondary wall synthesis phase of cotton fibre development.

Acknowledgements We are grateful to the Cotton Research and Development Corporation, Narrabri, New South Wales, Australia for generously supporting this work. These experiments comply with current Australian statutory requirements.

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidmann JG, Smith JA, Struhl K (1987) Current protocols in molecular biology. John Wiley and Sons, New York

Basra AS, Malik CP (1984) Development of the cotton fibre. Int Rev Cytol 89:65–112

Bernhard WR, Thoma S, Botella J, Somerville CR (1991) Isolation of a cDNA clone for spinach lipid transfer protein and evidence that the protein is synthesised by the secretory pathway. Plant Physiol 95:164–170

Brown CM, Stockwell PA, Trotman CNA, Tate WP (1990) Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. Nucleic Acids Res 18:6339–6345

Dellaporta SL, Wood J, Hicks JB (1985) Maize DNA miniprep. In: Molecular biology of plants: a laboratory course manual. Cold Spring Harbour Laboratory, Cold Spring Harbor, New York, pp 36–37

Delmer DP, Pear JR, Andrawis A, Stalker DM (1995) Genes encoding small GTP-binding proteins analogous to mammalian rac are preferentially expressed in developing cotton fibres. Mol Gen Genet 248:43–51

Dixon DC, Seagull RW, Triplett BA (1994) Changes in the accumulation of α and β-tubulin isotypes during cotton fibre development. Plant Physiol 105:1347–1353

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology and evolution of *Gossypium*. Adv Genet 23: 271–375

Graves DA, Stewart JM (1988) Chronology of the differentiation of cotton (*Gossypium hirsutum* L.) fibre cells. Planta 175:254–258

Heijne G von (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem 133:17–21

Hirzmann J, Luo D, Hahnen J, Hobom G (1993) Determination of messenger RNA 5′-ends by reverse transcription of the cap structure. Nucleic Acids Res 21:3597–3598

Hodgson CP, Fisk RZ (1987) Hybridisation probe size control: optimised "oligolabelling". Nucleic Acids Res 15:6295

John ME (1995) Characterisation of a cotton (*Gossypium hirsutum* L.) fibre mRNA (Fb-B6). Plant Physiol 107: 1477–1478

John ME, Crow LJ (1992) Gene expression in cotton (*Gossypium hirsutum* L.) fibre: cloning of the mRNAs. Proc Natl Acad Sci USA 89:5769–5773

John ME, Keller G (1995) Characterisation of mRNA for a proline-rich protein of cotton fibre. Plant Physiol 108: 669–676

John ME, Keller G (1996) Metabolic pathway engineering in cotton: biosynthesis of polyhydroxybutyrate in fibre cells. Proc Natl Acad Sci USA 93:12768–12773

John ME, Stewart JM (1992) Genes for jeans: biotechnological advances in cotton. Trends Biotech 10:165–170

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 15:6643–6653

Kader J-C (1996) Lipid transfer proteins in plants. Annu Rev Plant Physiol 47:627–654

Ma D-P, Tan H, Si Y, Creech RG, Jenkins JN (1995) Differential expression of a lipid transfer protein gene in cotton fibre. Biochim Biophys Acta 1257:81–84

Ma D-P, Liu H-C, Tan H, Creech RG, Jenkins JN, Chang Y-F (1997) Cloning and characterisation of a lipid transfer protein gene specifically expressed in fibre cells. Biochim Biophys Acta 1344:111–114

Masuta C, Furuno M, Tanaka H, Yamada M, Koiwai A (1992) Molecular cloning of a cDNA clone for tobacco lipid transfer protein and expression of the functional protein in *Escherichia coli*. FEBS Lett 311:119–123

Murray EE, Lotzer J, Eberle M (1989) Codon usage in plant genes. Nucleic Acids Res 17 : 477–497

Nakamura Y, Wada K, Wada Y, Doi H, Kanaya S, Gojobori T, Ikemura T (1996) Codon usage tabulated from the international DNA sequence databases. Nucleic Acids Res 24 : 214–215

Nothwehr SF, Gordon JI (1990) Targeting of proteins into the eukaryotic secretory pathway: signal peptide structure/function relationships. Bioessays 12 : 479–484

Orford SJ, Timmis JN (1997) Abundant mRNAs specific to the developing cotton fibre. Theor Appl Genet 94 : 909–918

Orford SJ, Timmis JN (1998) Specific expression of an expansin gene during elongation of cotton fibres. Biochim Biophys Acta 1398 : 342–346

Pyee J, Yu H, Kolattukudy PE (1994) Identification of a lipid transfer protein as the major protein in the surface wax of broccoli (*Brassica oleracea*) leaves. Arch Biochem Biophys 311 : 460–468

Rinehart JA, Petersen MW, John ME (1996) Tissue-specific and developmental regulation of cotton gene FbL2A. Plant Physiol 112 : 1331–1341

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York

Shimizu Y, Aotsuka S, Hasegawa O, Kawada T, Sakuno T, Sakai F, Hayashi T (1997) Changes in levels of mRNAs for cell wall-related enzymes in growing cotton fibre cells. Plant Cell Physiol 38 : 375–378

Song P, Allen RD (1997) Identification of a cotton fibre-specific acyl carrier protein cDNA by differential display. Biochim Biophys Acta 1351 : 305–312

Sterk P, Booij H, Schellekens GA, Van Kammen A, de Vries SC (1991) Cell-specific expression of the carrot EP2 lipid transfer protein gene. Plant Cell 3 : 907–921

Thoma S, Kaneko Y, Somerville C (1993) A non-specific lipid transfer protein from *Arabidopsis* is a cell wall protein. Plant Jour 3 : 427–436

Torres-Schumann S, Godoy JA, Pintor-Toro JA (1992) A probable lipid transfer protein gene in induced by NaCl in stems of tomato plants. Plant Mol Biol 18 : 749–757

Wan C-Y, 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