Analysis of Gene Expression During Cotton Fibre Development

by

Sharon Josephine Orford, B.Sc.(Hons)

A thesis submitted for the degree of

Doctor of Philosophy

in

The University of Adelaide

Department of Genetics

(Faculty of Science)

May, 1996

It's been a bloody long time coming,

but geez it's been worth it.

Russell Ebert, 1977

TABLE OF CONTENTS

CHAPTER 4: Abundant mRNAs Specific to the Developing Cotton Fibre

ABSTRACT

Cotton fibres represent a coordinated process of gene regulation within a single cell type and are especially suitable for studies of cellular and developmental events. In addition, the commercial desirability of long fibres has stimulated studies of the factors involved in controlling the extent of fibre growth. The aims of this project were to isolate and characterise cDNA clones of mRNAs which are specific to, or important in, cotton fibre development and to elucidate ribosomal gene expression and its role in fibre development.

Screening and sequencing of several low abundance cDNA clones from an initial cDNA library constructed from total RNA led to the isolation of a cDNA which is likely to encode a member of the superfamily of translation elongation factor 1α (EF-1α) proteins. EF-1α is universally conserved and is abundant in tissues which are active in protein synthesis. Southern analysis suggests that EF-1α is encoded in the cotton genome by a family of related genes, six of which were isolated within genomic clones.

Differential screening of a 13 DPA cotton fibre cDNA library constructed from poly(A)$^+$ RNA resulted in the identification of six putative fibre-specific messages of which five (pFS1, pFS3, pFS6, pFS17 and pFS18) showed preferential hybridisation to transcripts in fibre RNA. The mRNAs corresponding to these cDNAs exhibited differing patterns of temporal accumulation during fibre development.

The nucleotide sequences of three of the fibre-specific clones were similar to known sequences. The sequence of pFS1 showed similarity to a previously reported fibre-specific cDNA, encoding a protein of unknown function. pFS17 encodes a member of a class of well-characterised proline-rich structural proteins (PRPs) which are present in the walls of plant cells and a role for PRPs during elongation of the fibre cell is envisaged.

Messenger RNA corresponding to the remaining clone, pFS6, was the most abundant fibre-specific transcript. The pFS6 nucleotide sequence and its derived amino acid sequence

showed substantial similarity to phospholipid transfer proteins (LTPs), a class of plant proteins which has been implicated in the biogenesis of cellular membranes or in the deposition of cutin. A gene family of at least six members potentially encodes LTP-like mRNAs in *G. hirsutum*, of which two were isolated in a single 18 kb genomic clone. Sequence analysis of this clone suggested that neither is likely to encode the pFS6 group of fibre-specific transcripts. Screening of nucleotide databases was uninformative for the remaining two clones, pFS18 and pFS3.

Previous studies suggested that ribosomal RNA metabolism is related to final fibre length in cotton. The nucleolar size in fibres from three *G. hirsutum* varieties, differing in their final fibre lengths, was measured at early stages of growth. The nucleolar growth profiles differed between the three varieties, confirming previous work. One of several levels at which accumulation of rRNA may be controlled is by quantitative variation in the number of copies of the rRNA gene. The rDNA repeat unit from *G. hirsutum* var. Deltapine 90 was cloned and used to estimate the number of rRNA gene copies in six cotton varieties. Significant differences were observed between some of the varieties, but these did not correlate with final fibre length.

Results from this work clearly have commercial potential. Manipulation of the structure of fibre-specific genes provides exciting prospects for the modification of cotton fibre characteristics in transgenic plants. Alternatively, promoters isolated as a result of this study could be used to control the expression of heterologous genes specifically within the cotton fibre.

DECLARATION

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Signed:

Sharon Orford

27th May, 1996

ACKNOWLEDGEMENTS

Firstly, I wish to sincerely thank my supervisor, Assoc. Prof. Jeremy Timmis for his constant advice, encouragement, enthusiasm and friendship during the course of this work and his undying belief in the project and in myself.

I am grateful for the generosity of the Cotton Research and Development Corporation (Narrabri , NSW) who, in addition to project funding, supported me financially in the form of an Extended Studentship and a Travel Grant. Without their support this work would not have been possible.

There are several people who have contributed to this project on a professional level. I am indebted to Dr Danny Llewellyn (CSIRO Division of Plant Industry, Canberra) and Prof. J McD Stewart (Department of Agronomy, University of Arkansas, USA) for helpful discussions at the beginning of this project and to Dr Thea Wilkins for providing the RNA extraction protocol prior to its publication. Thankyou to Brian More at the Adelaide Botanic Gardens who helped in the location of the Australian native cotton species and to Dr Curt Brubaker who supplied material from the ancestral cotton species. I am indebted to the cotton breeders Peter Reid, Greg Constable and especially Norm Thomson, all of whom I have spoken to during the course of this study and who supplied various cotton seeds. Thankyou to Richard MacCallum for supplying seeds of *G. barbadense* and to Bill McDonnell for sending cotton seeds utilised in the early stages of the project. I am grateful also to Graham Teusner (Department of Metallurgy, The Levels) for the use of his microscope and micrometer.

I wish to thank all the occupants of the JT lab, both past and present, for making it an enjoyable and interesting place to work. Thankyou to Michael Ayliffe, Jamie Brady and Angela Binns for assistance in the initial stages of the project and for helping me find my feet in the lab after a break from the chaos of Honours. More recently, I wish to thank Tricia Franks for her copious and excellent advice and for rescuing me from all sorts of crises. Thanks Franks.

Special thanks also to Tamzin Donald for being such a great lab and conference buddy.

Thankyou to all the members of the Genetics Department, both past and present, for their direct and indirect contributions to this work. I am grateful to Marelle Smith and Robert Shroff for their invaluable help in computer matters and for being all-round good guys. Many thanks also to Paul Moir for putting up with my constant pestering, especially during these last days. All these people, and others, were responsible for creating an enjoyable and stimulating work environment.

Finally and most importantly, I would like to thank my friends and family, especially my Mum and Dad, for their love, interest and support throughout this project and for tolerating my scientific eccentricities. Thankyou also to my other families, the Harveys and Murphys. Particular thanks to Rachael and Richard Murphy for their continuing friendship and numerous conference jaunts and spending frenzies.

I wish to dedicate this thesis to John Harvey, for his friendship, understanding, encouragement and tolerance of both me and this thesis. *Dum vivimus vivamus.*

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	adenine
[α–^{32}P]dATP	alpha-labelled deoxyadenosine triphosphate
[α–^{32}P]dCTP	alpha-labelled deoxycytosine triphosphate
AGP	arabinogalactan protein
bp, kb	base pairs, kilobase pairs
C	constant, the DNA amount in an unreplicated haploid genome
C	cytosine
CaMV	cauliflower mosaic virus
cDNA	deoxyribonucleic acid complementary to ribonucleic acid
cm, μm	centimetre(s), micrometre(s)
cpm	counts per minute
CsCl	caesium chloride
C-terminal	carboxy-terminal
cv	cultivar
Da, kDa	Dalton(s), kiloDaltons
dATP	2'-deoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytosine-5'-triphosphate
DEPC	diethyl pyrocarbonate
dGTP	2'-deoxyguanosine-5'-triphosphate
dH_2O	distilled water
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNase I	deoxyribonuclease I
DPA	days post-anthesis
DTT	dithiothreitol
dTTP	2'-deoxythymidine-5'-triphosphate
EDTA	ethylenediaminetetraacetic acid
EtBr	ethidium bromide
g	force of gravity
(-40)	forward sequencing primer
G	guanine
g, mg, μg, ηg, ρg	gram(s), milligram(s), microgram(s), nanogram(s), picogram(s)
[γ–^{32}P]dATP	gamma-labelled deoxyadenosine triphosphate
GRP	glycine-rich protein
HRGP	hydroxyproline-rich glycoprotein

IPTG	isopropyl-β-D-thiogalactoside
λ	bacteriophage Lambda
L, ml, µl	litre(s), millilitre(s), microlitre(s)
LTP	lipid transfer protein
M	moles per litre
mA	milliAmperes
mCi	microCuries
min, hr	minute(s), hour(s)
MOPS	3-(N-Morpholino)propanesulphonic acid
Mr	molecular weight
mRNA	messenger ribonucleic acid
N-terminal	amino-terminal
NOR	nucleolar organising region
nt	nucleotide(s)
NTS	non-transcribed spacer
^{o}C	degrees Celsius
OD_{260}	optical density at 260 ηm
OD_{600}	optical density at 600 ηm
ORF	open reading frame
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
$poly(A)^{+}$	polyadenylated
$poly(A)^{-}$	non-polyadenylated
PRP	proline-rich protein
PVP	polyvinylpyrrolidine
5'-RACE	rapid amplification of 5' cDNA ends
rDNA	ribosomal deoxyribonucleic acid
RNA	ribonucleic acid
RNaseA	ribonuclease A
rpm	revolutions per minute
rRNA	ribosomal RNA
RSP	reverse sequencing primer
SDS	sodium dodecyl sulphate
SSC	saline sodium citrate
T	thymidine

Taq	*Thermus aquaticus*
TEMED	NNN'N'-Tetramethylethylenediamine
TFR	transferrin receptor (human)
Tris	Tris[hydroxymethyl]amino methane
U	unit(s) of enzyme
UTR	untranscribed region
UV	ultraviolet
V	Volts
X-gal	5'-bromo-4-chloro-3-indoyl-β-D-galactopyranoside
% (v/v)	percent volume per volume
% (w/v)	percent weight per volume

CHAPTER 1

Introduction

1.1 COTTON AS A CROP PLANT

1.1.1 The Cotton Industry

Cotton is grown as an annual field crop in over 90 countries in the temperate and tropical regions of the world. An increasing world population and a resurgence in the desirability of natural fibres for clothing has increased the demand for cotton and the cotton industry contributes $900 million to the Australian economy in a non-drought year.

The commercially important product of cotton is the seed coat hairs, commonly referred to as lint or fibre. After harvesting, the lint is separated from the seed by processing in a cotton gin, which also removes foreign matter while preserving the inherent properties of the fibre (Baker and Griffin, 1984). Although cotton is grown mostly for its fibre, the seeds are used as a source of edible oil and the seed meal is used as a protein-rich feed for livestock (Cherry and Leffler, 1984).

1.1.2 Classification and Evolution of Cultivated *Gossypium* Species

The cotton fibre of commerce is a product of Malvaceous plants of the genus *Gossypium*, which contains over 40 species, both diploid and tetraploid (Table 1.1). The many species of wild cotton are lintless or produce short fibres which are unsuitable for spinning. Cotton fibres are classed as trichomes and probably evolved as a dispersal mechanism for the cotton seed, mediated by wind or water (van der Pijl, 1982; Augspurger, 1986).

Evidence for the beginnings of cotton agriculture in the Near East date back 7000 or 8000 years and fabrics woven from cotton are known from 900BC (MacNeish, 1964). Cotton was domesticated independently in the New and in the Old World, with different species involved in the different centres of domestication. *Gossypium arboreum* L. and *G. herbaceum* L., the Asiatic cottons, are both diploid and native to the Old World (Lee, 1984). *Gossypium barbadense* L. and *G. hirsutum* L. are both allotetraploids, evolved in the New World and are agronomically superior to the diploids, having adapted to growth in a wide range of

Table 1.1: The species of *Gossypium*

Species	Genomic group[a]	Distribution
Diploids ($2n$ = 26)		
G. herbaceum L.	A_1	Old World cultigen
G. h. var. *africanum* (Watt) Mauer	A_1	Africa
G. arboreum L.	A_2	Old World cultigen
G. anomalum Wawr. and Peyr.	B_1	Africa
G. triphyllum (Harv. and Sand.) Hochr.	B_2	Africa
G. capitis-viridis Mauer	B_3	Cape Verde Islands
G. sturtianum J. H. Willis	C_1	Australia
G. sturtianum var. *nandewarense* (Derera) Fryx.	C_{1-n}	Australia
G. robinsonii F. Muell.	C_2	Australia
G. australe F. Muell.	—	Australia
G. costulatum Tod.	—	Australia
G. cunninghamii Tod.	—	Australia
G. nelsonii Fryx.	—	Australia
G. pilosum Fryx.	—	Australia
G. populifolium (Benth.) Tod.	—	Australia
G. pulchellum (C.A. Gardn.) Fryx.	—	Australia
G. thurberi Tod.	D_1	Mexico, Arizona
G. armourianum Kearn.	D_{2-1}	Mexico
G. harknessii Brandg.	D_{2-2}	Mexico
G. klotzschianum Anderss.	D_{3-k}	Galapagos Islands
G. davidsonii Kell.	D_{3-d}	Mexico
G. aridum (Rose & Standl.) Skov.	D_4	Mexico
G. raimondii Ulbr.	D_5	Peru
G. gossypioides (Ulbr.) Standl.	D_6	Mexico
G. lobatum Gentry	D_7	Mexico
G. laxum Phillips	D_8	Mexico
G. trilobum (DC.) Skov.	D_9	Mexico
G. turneri Fryx.	—	Mexico
G. stocksii Mast.ex. Hook.	E_1	Arabia
G. somalense (Gürke) Hutch.	E_2	Africa
G. areysianum (Defl.) Hutch.	E_3	Arabia
G. incanum (Schwartz) Hillc.	E_4	Arabia
G. ellenbeckii (Gürke) Mauer	—	Africa
G. longicalyx Hutch. and Lee	F_1	Africa
G. bickii Prokh.	G_1	Australia
Allotetraploids ($2n = 4x = 52$)		
G. hirsutum L.	$(AD)_1$	Central America
G. barbadense L.	$(AD)_2$	South America
G. tomentosum Nutt. ex Seem.	$(AD)_3$	Hawaii
G. mustelinum Miers ex Watt	$(AD)_4$	Brazil
G. darwinii Watt	$(AD)_5$	Galapagos Islands
G. lanceolatum Tod.[b]	(AD)	Mexico

[a] A dash (—) indicates that the genome designation has not been determined.
[b] Species status of *G. lanceolatum* needs experimental verification.

From Endrizzi *et al.* (1985)

environments as well as producing higher yields and longer fibres (Hutchinson, 1959; Horsfall *et al.*, 1972). *G. hirsutum*, or Upland cotton, contributes about 90% of world cotton production. *G. barbadense* cultivars include Sea Island, Egyptian and Pima varieties and are grown for their uniquely long, fine and strong fibre which supplies about 8% of world production of fibre (Lee, 1984).

The diploid species ($2n$=26) have been divided into ~~six~~ seven genomic groups, A, B, C, D, E ~~and~~ and G F, based on cytotaxonomic relationships (Table 1.1) and a total of five tetraploid ($2n$=$4x$=52) species are recognised. The allotetraploid (*AD* genome) of *Gossypium hirsutum* originated following a polyploidisation event which occurred 1.1-1.9 million years ago between two diploids (Fryxell, 1979; Reinisch *et al.*, 1994). Using mitochondrial DNA comparisons, Wendel (1989) concluded that the maternal or African (*A* genome) donor is similar to present-day *G. herbaceum* and *G. arboreum*, and it is believed that the ancestral American (*D* genome) parent is *G. raimondii*, a native Peruvian plant (Stephens, 1944; Endrizzi *et al.*, 1985).

1.1.3 Cotton Fibre Quality Parameters and Determination

The quality of fibres is important to the spinning and weaving industry and determines both the subsequent use of the fibre and the price paid for the crop. Several characters are used to assess quality. Grade classification is based on appearance and incorporates fibre colour and cleanliness (Perkins *et al.*, 1984). Staple length is the average length of the fibres on a seed, with five gradings from "short" staple Asian cottons (less than 21 mm) to "extra long" staple Egyptian and Sea Island cottons (35 mm and longer). Upland cotton cultivars are classified as "medium" (22-25 mm) and "medium-long" (25-28 mm) staple.

Fibre maturity, largely determined by the degree of secondary wall thickening which occurs before harvest, is of importance because immature fibres are not twisted and do not cling together when the fibre is spun (Basra and Malik, 1984). Fine fibres with high tensile strength is desirable because they are less likely to break during processing and produce strong

yarn. Fibre fineness is measured in combination with fibre maturity to give a micronaire reading, determined by an airflow measurement. Good quality fibre therefore consists of long, fine and strong fibres, but it is the length of the cotton fibre which largely determines the quality of the resultant thread.

1.1.4 Environmental Effects on Fibre Quality

High variability in desirable fibre properties often compromise the quality of both cotton yarn and fabric. Fibre characteristics, in particular length and strength, are greatly influenced by environmental factors and environment-genotype interactions (O'Kelley and Carr, 1953; Gipson and Joham, 1969; Gipson and Ray, 1969). Cool temperatures, for example, reduce the rate of fibre elongation (Gipson, 1986; Thaker *et al.*, 1989) and hinder the deposition of cellulose in the secondary cell wall of the fibre (Gipson, 1986; Haigler *et al.*, 1991). Soil compaction (Grimes *et al.*, 1975) and moisture availability (Sturkie, 1934), together with efficiency in the interception of photosynthetic light (Kasperbauer, 1994) also influence lint length.

1.1.5 Breeding for Improved Cotton Varieties

The modification of cotton is directed at improving both agronomic efficiency and product quality (Meredith *et al.*, 1991). Trends in cultivar improvement include onset of blooming, boll retention, boll maturation period, fibre and seed properties, environmental stress tolerance, boll size and prolificacy and pest resistance. The predominant method for cultivar improvement is hybridisation within and among types followed by selection for desirable characteristics. Heterotic responses in fibre properties and other traits of hybrids are well-documented (Thomson, 1971; Thomson and Luckette, 1988; Percy and Turcotte, 1992) and the large degree of intra-varietal variability in cotton has also been utilised as a basis for selection (Thomson, 1973a; Thomson, 1973b; Thomson, 1973c). Several mutant traits such as a developmental mutant which confers the preferred okra-shaped leaf phenotype (Hammond,

1941) have been incorporated into commercial cotton cultivars. Mutant stocks of *G. barbadense* and *G. hirsutum* are maintained in germplasm collections in the USA (Endrizzi *et al.*, 1984) and are available for exploitation in cotton improvement programs.

The majority of improved cotton cultivars currently in commercial use in Australia were bred at the Cotton Research Unit of the Commonwealth Scientific and Industrial Research Organisation at Narrabri in New South Wales. The Siokra cultivars incorporate the okra leaf mutation together with high yield, early maturation, good fibre quality and bacterial blight resistance traits together with suitability to Australian conditions, and compare favourably with American cotton cultivars (Thomson *et al.*, 1987; Colton, 1991). Cultivars with improved tolerance to the fungal wilt disease caused by *Verticillum dahliae* have also been produced using conventional breeding programs (Reid and Thomson, 1993).

There are several drawbacks in the use of conventional crossing methods in cotton. Hand-crossing of the largely self-pollinating plants is labour-intensive and the difficulty in producing F1 seed has limited the use of heterosis. Although the greater than 30 wild species of *Gossypium* (Table 1.1) present a large range of variation in many characters, genetic improvement through interspecific hybridisation is hampered by incompatibility barriers. Most characters for improving cotton are inherited as quantitative traits (Meredith, 1984), many of which have deleterious effects on fibre quality and lint yield (Meredith *et al.*, 1996). Fibre characters are generally heritable and thus readily selectable (Miller and Rawlings, 1967), but desirable fibre properties such as length and strength are often associated with low yield.

1.1.6 Transformation of Cotton

The advent of recombinant DNA technology and plant transformation opens new avenues for the production of improved cotton cultivars as it allows the transfer of genes conceivably from any source. Transformation of genes from other organisms will not only enhance existing fibre properties but may add new properties, complementing existing breeding

programs.

Agrobacterium-mediated transformation has been used to introduce genes into cotton for some years (Firoozabady *et al.*, 1987; Umbeck *et al.*, 1987) and the success rate improved through development of a reliable protocol for regeneration of plants from cell cultures (Trolinder and Goodin, 1987). The techniques have been successfully applied to Australian commercial cultivars (Cousins *et al.*, 1991). More recently, particle bombardment (Finer and McMullen, 1990; McCabe and Martinell, 1993) has been applied to the transformation of cotton and offers cultivar independence, faster plant regeneration and opportunities for multiple gene transfer (John and Stewart, 1992). An insect control protein has been successfully transformed into both Australian and American commercial cotton cultivars with the resultant transformants showing stable inheritance of the gene and protection from damage by cotton bollworm (Perlak *et al.*, 1990).

1.2 The Cotton Fibre as an Experimental System

In addition to its commercial importance, the cotton fibre has several attributes that recommend it as an experimental system for studying plant cell differentiation. Cotton fibres develop synchronously in the boll, providing a relatively homogeneous population of cells which can be readily detached from the seeds. Sufficient material of a given developmental stage can be obtained for biochemical and molecular analysis. Moreover, single-celled cotton fibres undergo a striking amount of elongation during their development, allowing investigation of the processes of cell expansion and cell wall synthesis free from the complications of cell division. Since the secondary cell wall consists almost entirely of cellulose, the cotton fibre also provides an excellent system for the elucidation of biochemical pathways of cellulose biosynthesis.

In addition, the system is amenable to *in vitro* culture techniques. Ovules in a defined culture medium undergo normal morphogenesis including fibre production (Beasley, 1977) and

fibre development of ovules grown *in vitro* is similar that in field-grown ovules (Meinert and Delmer, 1977; Triplett and Timpa, 1995). Cotton ovule culture has been used primarily to study the nutritional and hormonal requirements for fibre growth and development (Beasley, 1973; Beasley and Ting, 1973; Beasley and Ting, 1974; Beasley, 1977).

1.3 Growth and Development of Cotton Fibres

Cotton fibres originate as epidermal outgrowths of the ovule and begin to differentiate at about the time of anthesis (Figure 1.1; Figure 1.2a, b). Numerous investigations at both the ultrastructural and biochemical levels have shown that fibre cells develop in four distinct but overlapping phases. An initiation period is followed by extensive cell elongation (primary wall synthesis) which continues for a few weeks. Secondary wall synthesis, which begins 16 to 19 days post-anthesis (DPA), results in the deposition of almost pure cellulose and overall wall thickening and continues until the cotton fibre is mature (Basra and Malik, 1984).

1.3.1 The Cotton Flower

Cotton plants begin to flower approximately two months after seed germination (Figure 1.3a, b). The cotton flower comprises the reproductive organs and five cream-coloured petals surrounded by three large serrated bracts (Figure 1.2a, b) The ovary consists of three to five fused carpels, each containing eight to twelve ovules from which the fibre cells grow. Cotton flowers aged one day post-anthesis (1 DPA) appear pink due to a pigmentation change which results from production of ethylene (O'Neill *et al.*, 1993; Figure 1.2c).

1.3.2 Initiation of Fibre Cells

The spherical expansion of certain epidermal cells above neighbouring cells on the day of anthesis characterises the beginning of fibre growth (Stewart, 1975). Fibres first appear at the crest of the funiculus (Figure 1.1a) and although fibre development within a boll is widely considered to be synchronous, fibre initials are delayed for a few hours at the chalazal region

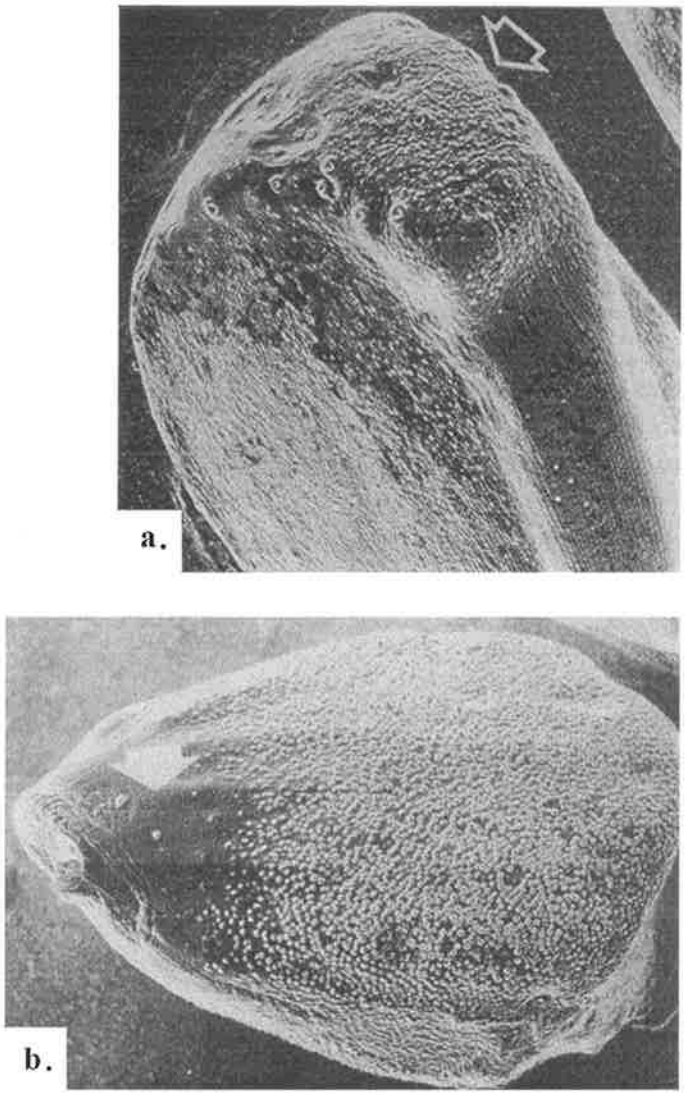


Figure 1.1: Cotton ovules at anthesis

Electron micrographs of cotton ovules on the day of anthesis, showing site and sequence of fibre cell initiation on the surface. From Stewart (1975).

a. Crest of funiculus (arrowed) where fibre initials first appear. The funiculus is adjacent to the chalazal end of the ovule and is the point of attachment of the developing ovule to the placenta.

b. Whole ovule showing fibres initiated in all areas except the micropylar end (arrowed).

and for up to four days at the micropylar region (Stewart, 1975; Figure 1.1b). The fibres that initiate elongation on about the day of anthesis are destined to become the lint of commerce whereas fibres which initiate in waves from about 4 to 12 days after anthesis (DPA) (Basra and Malik, 1984; Ryser, 1985) become short "fuzz" fibres or "linters", which have little commercial value and remain adhered to the seeds during ginning (Lang, 1938).

Although all epidermal cells have the potential to differentiate into fibre initials, only one out of every four epidermal cells elongates (Triplett and Timpa, 1995). The events determining whether a particular epidermal cell will differentiate into a cotton fibre are as yet unknown but initiation is apparently triggered by plant growth factors, as indicated by the effects of auxin and gibberellic acid on fibre production *in vitro* (Beasley and Ting, 1973; Beasley and Ting, 1974). Cells destined to be fibre cells can be distinguished from surrounding epidermal cells based on ultrastructure (Ramsey and Berlin, 1976a; Ramsey and Berlin, 1976b) and *in vitro* experiments have shown that the potential for fibre formation is present as early as 9 days preanthesis (Davidonis, 1989).

Several morphological alterations are associated with the early stages of fibre elongation. Formation of a large central vacuole occurs from the base of the fibre and occupies most of the cell volume by 2 DPA. A thin rim of cytoplasm separates the vacuolar membrane (tonoplast) from the cell wall during elongation, with the various organelles concentrated at the fibre tip. Fibre cells generally have an abundance of organelles including ribosomes, endoplasmic reticulum, Golgi bodies and mitochondria (Ramsey and Berlin, 1976a; Ramsey and Berlin, 1976b).

1.3.3 Elongation of Fibre Cells

The rate and duration of the elongation phase of fibre growth determines final fibre length (Kohel et al., 1974; Jasdanwala *et al.*, 1977; Naithani *et al.*, 1982). The single-celled trichomes expand at rates exceeding 2 mm per day to attain lengths ranging from 2 cm to 3.5

Figure 1.2: Morphology of a developing cotton boll

a. Cotton flower on the day of anthesis or flowering.

b. Cross-section of a cotton flower on the day of anthesis, to show location of ovules within the ovary. Adapted from Raven *et al.* (1986).

c. Cotton flower one day after flowering, showing pink pigmentation of the petals.

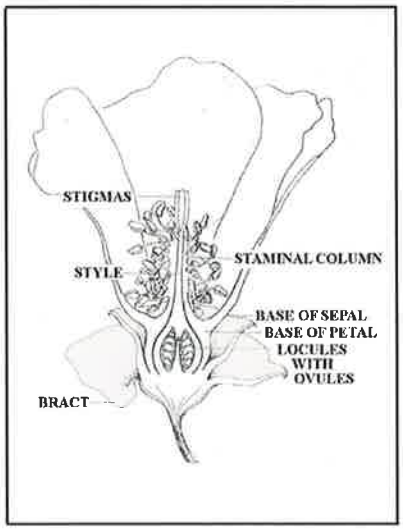
d. Developing cotton boll, five days after flowering.

e. Developing cotton boll, aged 20 days after flowering.

f. Mature cotton boll, aged approximately 50 days after flowering.

a.

b.


STIGMAS
STYLE
STAMINAL COLUMN
BASE OF SEPAL
BASE OF PETAL
LOCULES WITH OVULES
BRACT

c.

d.

e.

f.

Figure 1.3: Morphology of developing cotton plants

a. Development of Siokra 1-2 and Siokra L-22 plants under growth cabinet conditions.

i. Siokra 1-2 (right) and Siokra L-22 (left) plants two weeks after sowing of the seeds.

ii. Siokra 1-2 (right) and Siokra L-22 (left) plants six weeks after sowing of the seeds, showing extensive foliage growth.

iii. Siokra 1-2 plants four months after sowing, showing flowers and developing bolls.

iv. Siokra 1-2 plants five months after sowing, showing mature cotton bolls (arrowed) and loss of foliage from lower branches.

b. Development of an individual Siokra 1-2 cotton plant. The bar on the left represents a length of 20 cm.

i. A juvenile plant two weeks after sowing of the seeds.

ii. A Siokra 1-2 plant six weeks after sowing of the seeds, showing the okra leaf shape and extensive foliage growth.

iii. A Siokra 1-2 plant ten weeks after sowing of the seeds and after commencement of flowering.

iv. A Siokra 1-2 plant five months after sowing of the seeds, showing bolls and flowers at varying stages of development, including bolls approaching maturity (arrowed).

a. i

ii

iii

iv

b. i

ii

iii

iv

cm, with the maximal rate of expansion occurring between 6 and 12 DPA (Meinert and Delmer, 1977; Naithani *et al.*, 1981). Rapid growth of the boll is observed in the first 20 days after flowering (Figure 1.2d, e; Figure 1.3a, b), concomitant with the developing fibres and ovules within the carpels. Cotton fibres elongate throughout their entire length and not just at the tip (Ryser, 1977; Tiwari and Wilkins, 1995), and data from the incorporation of radioactively-labelled wall precursors suggest that primary cell wall formation also occurs along the entire fibre cell with some bias at the apex (O'Kelley, 1953). The elongation phase of fibre growth involves many different processes which are precisely controlled but cell wall biosynthesis is the major synthetic activity in cotton fibre cells and will be considered in some detail.

1.3.3a Primary Cell Walls of Plants

The primary cell wall which is deposited during fibre elongation is similar in composition to other dicotyledonous plants (Meinert and Delmer, 1977) and at 10 DPA consists of 35-50% cellulose plus approximately equal amounts of uronic acids, sugars, non-cellulosic glucans and proteins (Huwyler *et al.*, 1979; Timpa and Triplett, 1993). The matrix of primary cell walls is complex with hundreds, probably thousands of genes acting upon its formation and dynamics (Varner and Lin, 1989). Cell wall growth involves not only the synthesis of materials but their organised integration into the existing cell wall and in cotton fibres, the endoplasmic reticulum, Golgi complex and plasmalemma form a functionally integrated membrane system for the synthesis and transport of cell wall components (Ramsey and Berlin, 1976a; Ramsey and Berlin, 1976b). Current models for the typical angiosperm primary cell wall (Talbott and Ray, 1992; Carpita and Gibeaut, 1993) are based on three intermolecular networks. These networks, and the interactions between them, endow the wall with its mechanical properties and control whether, and at what rate, a plant cell can grow.

The first cell wall network is made up of cellulose microfibrils which are coated with an

array of polysaccharides, broadly classed as hemicelluloses and mainly consisting of xyloglucans (McQueen-Mason, 1995). Xyloglucan chains are hydrogen-bonded to the surface of the microfibrils and span neighbouring microfibrils, forming a continuous cohesive lattice throughout the wall. Direct microscopic evidence has been provided for the presence of xyloglucan cross-bridges between cellulose microfibrils (McCann *et al.*, 1990; Whitney *et al.*, 1995) and deposition of xyloglucan is limited to the elongation phase of cotton fibre development (Timpa and Triplett, 1993). This cellulose domain is embedded in a network of pectin, consisting of three polysaccharides, homogalacturonan, rhamnogalacturonan I and rhamnogalacturonan II, which form a gel-like pectin matrix. Structural proteins form a third interlocking network (reviewed in McCann and Roberts, 1994).

In addition to the extensive cell wall synthesis which occurs during elongation, a thin cuticle is deposited on the surface of the primary wall and stretches and thins as the fibre cell elongates (Flint, 1950; Willison and Brown, 1977; Ryser, 1985). Consisting of fats, waxes and resins, the cuticle is a characteristic of cells which are exposed to air at some stage in their life cycle and protects the living cell from desiccation (Kolattukudy, 1981).

1.3.3b Control of Cell Expansion

The control of fibre elongation is an important consideration in cotton bioengineering and there are many levels at which fibre growth can be controlled. Plant cell enlargement is tightly regulated by plant hormones, pH, light and other factors and is driven by turgor pressure, which in cotton fibre cells is largely generated by the influx and accumulation of potassium and malate in the enlarged central vacuole (Dhindsa *et al.*, 1975). Malate and K^+ levels fluctuate in correlation with growth rate of both *in vivo* and *in vitro* grown fibres, reaching a peak when growth rate is maximal (Dhindsa *et al.*, 1975). In addition, amounts of malate synthesising enzymes, K^+ and malate were found to be higher in elongating fibres of *G. hirsutum* (a long staple type) than in its short-fibred counterpart, *G. arboreum* (Basra and

Malik, 1983). The differential accumulation of other proteins involved in maintenance of turgor in the developing fibre cell also parallels the rate of elongation. For example, the expression of vacuolar H^+-ATPase subunit genes is induced specifically in tissues undergoing rapid cell expansion and altered patterns of expression are observed in cotton genetic mutants exhibiting aberrant trichome development (Joshi *et al.*, 1988; Wilkins and Wan, 1993; Hasenfratz *et al.*, 1995).

The growth (in area) of a cell wall requires the cellulose microfibrils to move apart or to slip relative to each other, and this requires breakage or lengthening of the polysaccharide linkages. "Wall-loosening" enzymes are thought to initiate turgor-driven cell expansion by biochemical modification and relaxation of the cell wall. Expansins, first described in cucumber hypocotyls (McQueen-Mason *et al.*, 1992), are a class of cell wall proteins which appear to loosen the wall via a non-hydrolytic mechanism that disrupts the non-covalent adhesion of cell wall matrix polysaccharides to each other or to cellulose microfibrils (Cosgrove, 1993; McQueen-Mason and Cosgrove., 1994; McQueen-Mason and Cosgrove, 1995). Endo-xyloglucan transferase (XET) is a recently identified transferase which can catalyse the molecular grafting reactions between xyloglucan molecules (Nishitana and Tominaga, 1992). Correlations have been found between XET activity and plant growth rate (de Silva *et al.*, 1994) and modification of wall matrix polysaccharides by enzymes such as hydrolases and XET may work synergistically to enhance expansin-mediated extension (Cosgrove and Durachko, 1994; Fry, 1994).

1.3.3c Hormonal Considerations

Auxin is involved in the enhancement of cell elongation and enlargement (Hong *et al.*, 1987). In all cotton cultivars, auxin levels in fibre cells peak at about the time of commencement of elongation and correlate positively with the final fibre length (Naithani *et al.*, 1982). Levels of auxin are regulated by enzymes such as *o*-diphenol oxidase and

peroxidases, the activities of which are considerably higher during the elongation phase in short staple cultivars than in their medium and long staple counterparts (Basra and Malik, 1981). In addition, auxin catabolism is low during the elongation phase of *G. hirsutum* fibres but increases four-fold during secondary thickening and this trend is reflected in the relative activity of IAA oxidising systems (Jasdanwala *et al.*, 1977).

The mechanism which underlies auxin action on the enhancement of cell elongation is not fully defined but one early manifestation of auxin treatment is the regulation of gene expression at the transcriptional level and an increase in the synthesis of some cell wall constituents (Varner and Lin, 1989). The application of growth substances or manipulation of pertinent enzyme systems may prove to be effective in increasing the number, length, strength and uniformity of cotton fibres (Basra and Malik, 1984) and is currently under investigation (John and Stewart, 1992; John, 1994).

1.3.4 The Transition from Elongation to Secondary Wall Deposition

At approximately 16 to 18 DPA, fibre elongation slows with the onset of secondary wall synthesis and the two phases overlap by several days. There seem to be varietal differences in cotton regarding the capacity for elongation after the beginning of secondary wall deposition and varying degrees of overlap between the phases of elongation and secondary wall synthesis have been reported by Benedict *et al.* (1973), Schubert *et al.* (1973), Jasdanwala *et al.* (1977) and Huwyler *et al.* (1979). Degree of overlap between primary and secondary wall synthesis and timing of secondary wall synthesis may correlate with fibre quality, particularly final fibre length (Benedict *et al.*, 1973; Beasley, 1979). Overlap tends to be greater in long staple cultivars than in short staple cultivars (Naithani *et al.*, 1982).

The signals which trigger fibre cells to enter the next phase of growth are unknown, but differences in the molecular profile of cotton fibre during the stages of elongation and secondary wall synthesis are substantial and onset of the latter phase is characterised by several

biochemical changes. For example, callose, a β-1,3-linked glucan, is normally found associated with wound response in plants but is synthesised transiently in cotton fibres at the onset of secondary wall synthesis (Waterkeyn, 1981; Ryser, 1985; Maltby *et al.*, 1979). Speculation as to the role of β-1,3-glucan polymers has ranged from cell wall turnover, a reserve function or metabolic intermediate (Waterkeyn, 1981). Callose synthase may share common subunits or be an altered form of cellulose synthase (Meier *et al.*, 1981; Hayashi *et al.*, 1987; Delmer *et al.*, 1991; Bulone *et al.*, 1995).

Similarly, a significant amount of silicon is found in the cell walls of developing cotton fibres, with maximum levels corresponding with initiation of secondary wall thickening. A higher concentration of silicon in cotton varieties with long fibres suggests a possible role for silicon in fibre elongation and/or initiation of secondary wall synthesis (Boylston et al, 1990). Cotton bolls accumulate a pectic fraction of carbohydrates, mainly glucose, fructose and sucrose during the elongation phase development (Conner *et al.*, 1972). Huwyler *et al.* (1979) found that the absolute amounts of fucrose, galactose, mannose, arabinose and uronic acids all reached a maxima at the transition between primary and secondary wall formation and thereafter decreased, implying that degradation of noncellulosic polysaccharides was occurring.

The transition between elongation and secondary wall deposition is also marked by changes in protein patterns, as seen by SDS-PAGE of *in vitro* translation products (Alexander *et al.*, 1983; John and Stewart, 1992). The auxin content of cotton fibres decreases sharply prior to the onset of secondary wall formation (Basra and Malik, 1984) and it may be depletion in the levels of auxin which triggers wall deposition (Jasdanwala *et al.*, 1977).

1.3.5 Secondary Wall Deposition

Secondary wall synthesis is characterised by the deposition of a homogeneous population of high molecular weight cellulose on the inner surface of the existing primary wall, at the expense of the cell lumen. Deposition of the secondary wall begins at approximately 16 DPA

and continues until 30-35 DPA, with maximal deposition occurring at 26-28 DPA. The secondary cell wall is less hydrated and more compact than the primary wall and differs substantially in chemical composition (Ward and Moo-Young, 1989).

In a mature fibre the secondary wall is about 94% cellulose (Young, 1986) and relatively thick, providing the extraordinary strength required of a textile fibre. Secondary wall characteristics, such as molecular weight distributions of cellulose polymers, vary between cotton cultivars (Timpa and Ramey, 1989) and the length of cellulose chains has a direct influence on the tensile strength of the fibre cell (Benedict *et al.*, 1994). The disposition of cellulose fibrils in the secondary walls of cotton fibre cells is unique. Whereas in most cells the fibrils are angled axially or spiral continuously in one direction, the spiral pattern of fibrils in cotton fibres intermittently reverses direction. The frequency of such reversals varies between cotton cultivars (Betrabet *et al.*, 1963). The cell wall is thinner at reversal points than in other parts of the fibre and it is at these points that mature fibres twist (Section 1.3.6), resulting in a spinnable fibre.

1.3.6 Maturation of Fibre Cells

The final phase in fibre development, that of maturation, occurs 40 to 50 DPA and culminates in boll opening which causes fibre cell death and desiccation (Figure 1.2f, Figure 1.3a, b). As they dry, the cylindrical fibre cells collapse to twisted, ribbon-like structures in a way that allows the spinning of fine, strong cotton threads. The angle of spiral twisting together with the degree of secondary wall thickening affects fibre strength (Basra and Malik, 1984).

1.4 The Biochemistry of Cotton Fibre Development

1.4.1 Cellulose in the Cotton Fibre Cell Wall

Chemically, cellulose is a linear β-1,4-linked glucan. Uridine diphosphoglucose

molecules are polymerised into long cellulosic chains which are hydrogen-bonded to one another to form insoluble crystalline bundles called microfibrils. The current model for cellulose synthesis, supported by evidence from *Acetobacter xylinum*, a cellulose-producing bacterium (Saxena *et al.*, 1994), is that synthesis of microfibrils occurs on the plasma membrane in association with large multi-enzyme complexes (Westafer and Brown, 1976; Willison and Brown, 1977).

The cellulose synthase complexes are mobile in the plane of the plasma membrane and are guided in their motion by plant cytoskeletal elements. Cotton fibres contain highly organised arrays of cytoskeletal microtubules which appear to shift orientation in a developmentally-controlled manner. Microscopic and immunocytochemical evidence has shown that orientation of microtubules mirrors that of the cellulose microfibrils (Westafer and Brown, 1976; Willison and Brown, 1977; Seagull, 1986; Seagull, 1992) even throughout spiral reversals (Yatsu and Jacks, 1981; Section 1.3.5). In addition, chemical disruption of the cytoskeleton not only disorganises microtubules in cotton fibre cells but also deranges normal cell wall microfibrillar orientation (Yatsu and Jacks, 1981) and results in significant decreases in fibre textile properties (Yatsu, 1983).

The orientation and arrangement of cellulose microfibrils in the cell wall has an important bearing on the physical properties of cotton fibres. Control points for cellulose deposition could occur at several levels including production, transport or uptake of substrate, provision of sufficient energy through respiration, or function of the endomembrane system and cytoskeletal elements. The organisation of cellulose microfibrils can be modified through interactions with a variety of wall components (Preston, 1979). In addition, extensive efforts have been invested in the isolation of cellulose synthase from higher plants. The main barrier to the identification of cellulose synthase in higher plants has been the inability to synthesise cellulose *in vitro* using cell-free systems. Use of *Acetobacter xylinum* as a model experimental system resulted in characterisation of a cellulose-synthase operon (Wong *et al.*, 1990; Saxena

et al., 1991; Saxena *et al.*, 1994; Standal *et al.*, 1994) and identification of the catalytic subunit of cellulose synthase (Lin *et al.*, 1990). Antibodies against *Acetobacter* cellulose synthase cross-react with proteins present in several higher plants including the leaves of cotton (Mayer *et al.*, 1991). In a series of articles, Brown and colleagues (Okuda *et al.*, 1993; Li and Brown, 1993) reported the synthesis of appreciable amounts of β-1,4-glucan *in vitro*. Subsequently, the catalytic subunits of both callose and cellulose synthase were identified from cotton fibres (Li *et al.*, 1993; Kudlicka *et al.*, 1995) although results from the experiments have been disputed (Delmer *et al.*, 1993; Brown *et al.*, 1994).

Studies of various cotton mutants showed that cellulose deposition during the elongation and secondary wall deposition phases of fibre development are under separate genetic control (Kohel *et al.*, 1974). Fibre length is limited to less than 5 mm in two fibre mutants, Ligon lintless-1 and Ligon lintless-2, but secondary wall formation appears to occur normally (Kohel *et al.*, 1992). The rate of formation of crystalline cellulose in the primary fibre cell walls of Ligon lintless-2 was shown to be 40% less than in wild-type fibres (Kohel *et al.*, 1993) and it is the aberrant cellulose deposition which presumably results in the dramatic reduction in the elongation of mutant fibres. Since fibre elongation and secondary wall deposition are not necessarily controlled by the same genetic factors (Kohel *et al.*, 1974), it may be possible to alter through genetic manipulation either elongation or dry weight increase of the fibres without appreciably changing the other.

1.4.1a Cellulose in the Primary Cell Wall

Cellulose microfibrils in the primary cell wall cellulose consist of a heterogeneous population of molecules with generally low molecular weight. In fibre initials, the cell wall contains cellulose microfibrils which are organised in a random pattern and therefore do not hamper the "ballooning" of the cells above the ovule epidermal layer (Seagull, 1993; Section 1.3.2). Soon after initiation, microfibrils are reorganised into a loose arrangement transverse to

the direction of growth and thereafter, are deposited in a shallow pitched helical pattern which spirals down the length of the fibre. The distinctly transverse pattern of microfibril deposition is responsible for controlling the unidirectional elongation of the fibre cell (Seagull, 1993). As elongation continues, the orientation of the microfibrils changes so that, by 10 DPA, the outermost (oldest) microfibril layers are oriented in the direction of growth (axially) while new microfibrils are still deposited transversely (Seagull, 1993; Timpa and Triplett, 1993).

1.3.1b Cellulose in the Secondary Cell Wall

Secondary cell wall synthesis is characterised by a sharp increase in the rate of cellulose synthesis at 16 DPA which peaks at 17 DPA and declines by 18 DPA (Meinert and Delmer, 1977). The bulk of the secondary wall is deposited as layers of parallel cellulose microfibrils, organised in helical arrays with increasing pitch (Seagull, 1993) such that the microfibrils closest to the lumen have an axial orientation. The first layer of the secondary cell wall, termed the winding layer, can be morphologically distinguished from the remainder of the secondary wall and its deposition is thought to occur during the transition between the elongation and secondary wall deposition phases of cotton fibre growth. Microfibrils in the winding layer are not as densely packed as microfibrils in the rest of the secondary cell wall and do not completely restrict cell expansion, whereas onset of deposition of subsequent highly organised layers of the secondary wall restricts elongation completely (Meinert and Delmer, 1977; Seagull, 1993).

The shift in orientation of cellulose microfibrils from the transverse pattern of the primary wall to a sharply helical pattern of deposition is paralleled by similar changes in orientation of the cortical microtubule network from transverse to spiral between 15 and 18 DPA. Both events may therefore be regulated upstream by a reorganisation of actin (Seagull, 1992; Seagull, 1993), perhaps mediated by plant growth substances. Gibberellins have been found to control the polarity of maize root cell growth through regulation of the orientation and

arrangement of cortical microtubules (Baluška *et al.*, 1993) and it is reasonable to assume that they play a similar role in the developing fibre cell.

1.4.2 Proteins Associated with the Development of the Fibre Cell

1.4.2a Proteins in Plant Cell Walls

Cell walls in plants contain only relatively small amounts of protein, with primary walls containing more protein than secondary walls (Lamport, 1965; Keller, 1993). Cell wall proteins consist of both structural proteins and enzymes. Five classes of structural cell wall proteins have been characterised to date. These are the extensins, or hydroxyproline-rich glycoproteins (HRGPs), the proline-rich proteins (PRPs), the glycine-rich proteins (GRPs), the arabinogalactan proteins (AGPs) and the lectins, which are restricted to the Solanaceae family (Cassab and Varner, 1988; Keller, 1993; Showalter, 1993). Each class of structural cell wall protein has a distinctive repeating motif and is encoded by a small multigene family, the members of which are regulated in a cell-specific as well as in a developmental fashion. Their precise role in cell wall architecture in unknown but HRGPs and PRPs are thought to modify the plant cell wall by covalent crosslinking of tyrosine residues to form an insoluble network (Fry, 1986; Cooper *et al.*, 1987; Cassab and Varner, 1988). This would harden and strengthen the wall against environmental stress and pathogen attack or wounding (Bradley *et al.*, 1992).

GRPs are primarily associated with lignified plant cell walls (Keller *et al.*, 1988; Ye and Varner, 1991), which suggests that they influence the polymerisation of lignin or direct the sites of lignin deposition. Unlike the classes of protein described above, the arabinogalactan proteins (AGPs) and lectins are soluble wall components. AGPs are a complex group of molecules which are found in almost all tissues of higher plants and are probably non-structural components of the matrix (Marcus *et al.*, 1991). AGPs typically comprise only 2 to 10% protein by weight and increased concentration of AGPs is associated with maturation of the stigma, suggestive of a role in pollen incompatibility. The selective activation of genes

encoding specific structural proteins such as HGRPs and PRPs provides a mechanism for precise control of cell wall architecture during cellular differentiation and elongation.

Various enzymes taking part in the formation and modification of cell walls have been identified from a large number of plants. Enzymes such as peroxidases, phosphatases, invertases, glucanases, arabinosidases, proteases, pectinesterases and galactosidases as well as extensins and XET have been characterised. They have been implicated in defence and cell maturation as well as presumed roles in the degradation and turnover of cell constituents (Cassab and Varner, 1988; Varner and Lin, 1989). The best characterised cell wall glycoenzyme is the family of peroxidases, which have been shown to catalyse formation of crosslinks between macromolecules such as lignin, proteins and hemicelluloses (Cassab and Varner, 1988, Showalter and Varner, 1989). Peroxidases may therefore restrict cellular growth by rigidifying the cell wall and have been also been implicated in the regulation of auxin levels and response to microbial attack. In accordance with a role in cell expansion, peroxidases have altered patterns of expression in enhanced extension (*slender*) mutants of pea, tomato and barley (Jupe and Scott, 1989; Jupe and Scott, 1992; Schünmann *et al.*, 1994).

1.4.2b The Protein Profile of Developing Fibre Cells

There is considerable synthesis of proteins during cotton fibre development (Huwyler *et al.*, 1979). The percentage of proteins in the fibre cell wall peaks at the end of elongation, then steadily decreases as the fibre matures (Meinert and Delmer, 1977; Huwyler *et al.*, 1979). Investigation of the protein profile of developing cotton ovules and fibres showed that fibre development is characterised by at least two distinct protein populations, the first of which occurs around anthesis (Graves and Stewart, 1988a). The early proteins are believed to be involved in fibre initiation and establishment of the elongation phase. The second population of proteins appears around 10 DPA and is believed to have a role in the initiation of secondary wall deposition. This set of proteins is detectable beyond 20 DPA (Graves and Stewart,

1988a). Interestingly, the staining intensity of several specific proteins (Graves and Stewart, 1988a) increased in parallel with the reported pattern of cellulose deposition (Meinert and Delmer, 1977; Alexander *et al.*, 1983) and protein patterns differed between fibres and ovules of the same age, suggestive of tissue-specific proteins in both tissues (Graves and Stewart, 1988a).

1.4.2c Cytoskeletal Proteins in the Fibre Cell

Microtubules in the plant cytoskeleton are involved in the deposition of organised arrays of cellulose microfibrils in most, if not all, higher plant systems (Section 1.4.1). Tubulins, the main component of microtubules, are encoded by a well-characterised gene family, the members of which have constitutive or developmental stage-specific and/or tissue-specific expression (Silflow *et al.*, 1987). As a percentage of total protein fibre cells have greater amounts of tubulin than hypocotyls, roots, leaves or cotyledons (Dixon *et al.*, 1994). Consistent with increases in microtubule length and number that occur during fibre development, levels of α and β tubulin monomers increase in cotton fibres up to 20 DPA before reaching a plateau or decreasing slightly after the onset of secondary wall synthesis (Meinert and Delmer, 1977; Kloth, 1989; Seagull, 1992). Specific components of the cytoskeleton may play regulatory roles in defining fibre development and may provide subjects for the manipulation of not only fibre characteristics but general plant cell morphogenesis.

1.4.3 Supply of Metabolic Energy to the Fibre Cell

Sucrose is the principle form in which carbon and energy are supplied to non-photosynthetic plant cells such as the cotton fibre (Basra *et al.*, 1990; Jaquet *et al.*, 1982). Its immediate metabolism, regulated by the relative activities of acid invertase, alkaline invertase and sucrose synthase, is of fundamental significance in developing cotton fibres. A high level of invertase activity is localised in tissues with rapid growth (Gayler and Glasziou, 1972) and high activities of sucrose metabolising enzymes have been reported both in *in vitro* and *in vivo*

grown fibres (Beasley, 1977). In addition, the enzymes appear to have higher activities in longer than in shorter fibres (Basra *et al.*, 1990) and at least half of the total sucrose synthase is associated with the plasma membrane and may serve to channel carbon directly from sucrose to cellulose and/or callose synthesis (Amor *et al.*, 1995). Metabolic requirements during cotton fibre development may be partly met through interrelated operation of the glycolytic and pentose phosphate pathways (Basra and Malik, 1984). Higher activities of glycolytic and pentose pathway enzymes together with increased turnover of [^{14}C]glucose are observed in long staple as opposed to short staple cultivars (Basra, 1982).

1.5 Ribosomal RNA Genes and Development of the Cotton Fibre

1.5.1 Ribosomal RNA Genes in Higher Plants

The genomes of higher plants are characterised by very high percentages of repeat sequences and ribosomal RNA genes (rDNA) form a major class of tandemly repeated sequences. The genes which specify the rRNAs are organised in units which contain the regions encoding 18S, 5.8S and 25S ribosomal RNAs and an extended intergenic region sometimes termed the non-transcribed spacer or NTS (Figure 1.4). Ribosomal RNA genes occur in tandem arrays and are located at one or more chromosomal loci in a given plant genome, the so-called nucleolar organising regions (NORs). rRNA genes are transcribed as a single precursor or pre-rRNA which is subsequently processed via a complex series of steps into mature 18S, 5.8S and 25S subunits of the ribosome (Sollner-Webb and Tower, 1986).

The number of NORs and total number of rRNA genes vary between species, ranging from a few hundred to several thousand rDNA copies per genome (Ingle *et al.*, 1975; Rogers and Bendich, 1987a). General organisation of rRNA genes has been described in many species, both monocots and dicots (Vedel and Delseny, 1987) and the coding regions show substantial sequence conservation not only among plant species but among eukaryotes. In contrast, the

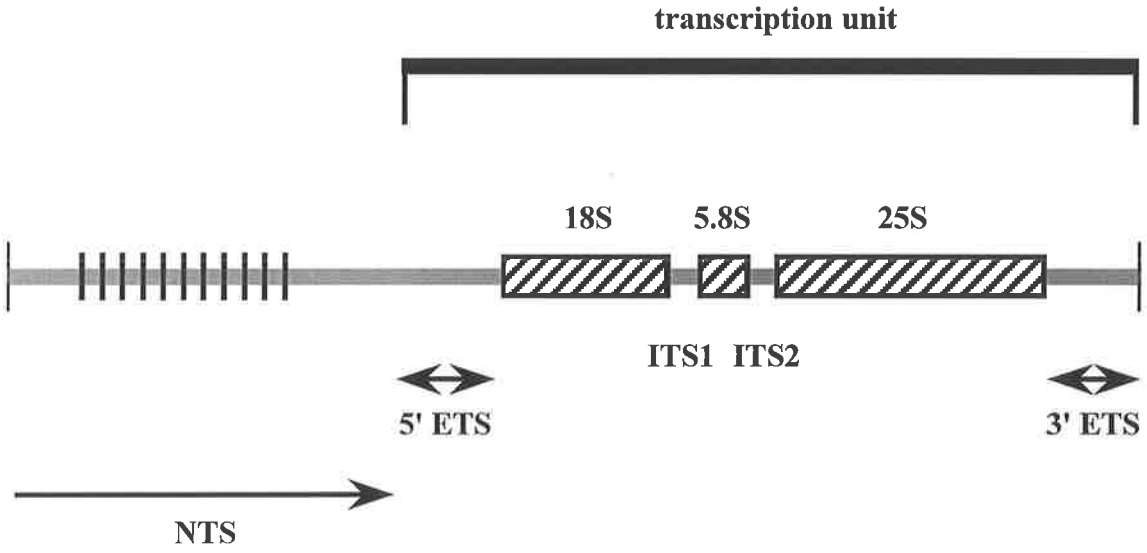


Figure 1.4: Schematic diagram of a single rRNA gene repeat unit

Map of an rDNA repeat unit showing positions of the 18S, 5.8S and 25S coding regions, which are separated by the internal transcribed spacers ITS1 and ITS2. External transcribed spacers (ETS) are indicated by arrows, as is the intergenic spacer or non-transcribed spacer, NTS with its repeat elements.

NTS and parts of the pre-rRNA removed during processing show considerable divergence between phylogenetic groups and between plant species.

In tetraploid cotton, three major 18S-25S loci were mapped by *in situ* hybridisation to chromosomes 9, 16 and 23, with an additional minor locus on chromosome 7 (Crane *et al.*, 1993). Sequence data from ITS1 and ITS2 regions (Wendel *et al.*, 1995; Figure 1.4) indicate that rDNA arrays are virtually homogeneous in several diploid and allotetraploid cottons. The 5S rRNA genes are organised in tandemly repeated clusters uninterrupted by other sequences but are separated physically from the rDNA clusters. (Hemleben and Grierson, 1978). In cotton the two 5S rDNA sites, which are different in size, have been localised by *in situ* hybridisation to regions near the centromeres of chromosomes 9 and 23 (Crane *et al.*, 1993).

1.5.2 The Nucleolus

The nucleolus is the site where major (18S-25S) ribosomal RNA genes are transcribed and the resultant 45S primary transcript is processed and assembled into preribosomal particles. The nucleolus therefore contains the transcribed rDNA and the different transcripts as well as the structural and enzymatic molecules involved in such activity, such as ribosomal proteins, RNA polymerase I, DNA topoisomerases, transcription factors, processing enzymes and the imported 5S rRNA.

Three distinct ultrastructural components are visible in electron micrographs of almost all nucleoli and the components appear to be dynamic and dependent on ultrastructural activity. Powerful *in situ* methodologies and immunoelectron microscopy, such as that carried out in pea root cells (Shaw *et al.*, 1995) and *Vicia faba* mitotic cells (Sato and Myoraku, 1994) have mapped different parts of the transcription, processing and assembly pathway to the nucleolar components. The various RNA processing events appear to be spatially highly organised, with the early steps of transcription and processing occurring in the centre of the nucleolus and later steps in its periphery (Risueño and Testillano, 1994).

Several lines of evidence suggest that the nucleolar volume is a measure of nucleolar activity and is closely related to the number of active rRNA genes at an NOR. Generally, plant cells engaged in rapid protein synthesis possess large nucleoli (reviewed in Peeters *et al.*, 1988) and active cells tend to have larger and more nucleoli than inactive cells. In addition, changes in number, size, morphology and shape of nucleoli have been noted in differentiating cells, depending on their metabolic and proliferative state (Ghosh, 1987).

1.5.3 Ribosomes and Nucleoli in Cotton Fibre Cells

Primordial fibre cells can be distinguished from the surrounding epidermal cells by several morphological features which include more numerous ribosomes and rough endoplasmic reticulum (Ramsey and Berlin, 1976a; Section 1.3.2). The nucleolus is also enlarged in prefibre cells (Ramsey and Berlin, 1976a) on the day of anthesis and further enlargement of the nucleus and nucleolus occurs by 2 DPA (Ramsey and Berlin, 1976b), with a peak at around 4-10 DPA depending on the cultivar. The nucleus maintains its size during further fibre development, while the size of the nucleolus declines gradually. No considerable changes in the nucleolus are observed after 20 DPA (Peeters *et al.*, 1987) and prominent nucleoli are not observed in the nuclei of older cotton fibres (Ramsey and Berlin, 1976b). Different cotton species are characterised by different nucleolar growth profiles (Peeters *et al.*, 1987) which suggests that nucleolar activity is under genetic control.

Results of statistical analysis of nucleolar size and final fibre dimensions have prompted the suggestion that large and small nucleoli belong to long and short fibres respectively (Peeters *et al.*, 1987; Peeters *et al.*, 1988). Nucleolar activity, expressed by the size of the nucleolus and perhaps its vacuolation, was less in fibres sampled from the micropylar region of the seed than in the longer fibres of the mid region (Peeters *et al.*, 1987). Changes in nucleolar size and vacuolation at early stages of fibre development may therefore influence the subsequent growth of the fibre.

Taken together, these observations suggest that developing cotton fibres possess a greater capacity for protein synthesis than the surrounding epidermal cells. The rapid and extensive period of elongation and primary cell wall synthesis by cotton fibre cells clearly requires substantial capacity for protein synthesis, especially in membrane formation necessary for the expanding plasma membrane and tonoplast (Ramsey and Berlin, 1976b). Ribosome synthesis necessary for cell elongation appears to occur early in the elongation period and declines as the fibre matures. It may be that the amount of ribosomes synthesised early in fibre development, up to approximately 8 DPA determines the subsequent rate and degree of elongation as well as the thickness of the fibre (DeLanghe *et al.*, 1978).

1.5.4 Control of Ribosomal RNA Gene Expression

Eukaryotic cells synthesise vast amounts of rRNA, accounting for over 80% of total steady-state cellular RNA. However, the total number of rRNA genes per cell is in excess of that required to sustain of ribosome synthesis, even in growing cells. Evidence for rRNA gene redundancy comes from the extensive variation in gene copy numbers in between and within individual plants (Rogers and Bendich, 1987a; Rogers and Bendich, 1987b), the often very large numbers of rRNA genes per cell, and the viability of plant mutants with a large percentage of their rRNA genes deleted (Flavell *et al.*, 1985). Birnstiel and coworkers (1971) have claimed that in *Xenopus* only 150 rDNA cistrons are necessary for protein synthesis required for normal maintenance of a cell. No association has been found ~~been found~~ between rDNA copy number and phenotype, vigour or rRNA content (Buescher *et al.*, 1984) and it has been suggested that the maintenance of repetitive DNA sequences, including rDNA, are maintained in plants to allow adaptation to unfavourable changes in their environment (Walbot and Cullis, 1985; Cullis, 1990).

Altered chromatin structure and demethylation are intimately related to rRNA gene expression (Flavell *et al.*, 1985; Flavell *et al.*, 1988; Scott *et al.*, 1984). Methylation studies

carried out on members of the Cucurbitaceae (Ingle *et al.*, 1975) and cucumber (Hemleben and Zentgraf, 1994) showed that at least 70% of the repeats are completely methylated. The small percentage of repeats which appears completely unmethylated are proposed to represent the actively transcribed portion of rDNA (Flavell *et al.*, 1988; Hemleben and Zentgraf, 1994). In general, very active nucleoli have a higher proportion of rRNA genes with unmethylated cytosine residues compared with less active nucleoli and the extent and pattern of methylation can alter during development (Flavell *et al.*, 1988). Interestingly, stimulation of rRNA synthesis is observed after treatment with an analogue of auxin (Guilfoyle *et al.*, 1975). This may have implications for cotton fibres, since levels of auxin are high during the elongation phase of development, correlating with increased numbers of ribosomes in fibre initials.

The number of rRNA genes engaged in transcription appears to be correlated to a cell's general metabolic and proliferative activity and the rate of rRNA synthesis can change enormously depending on the developmental and physiological state of the cells or the environmental conditions (reviewed in Sollner-Webb and Tower, 1986). Changes in rRNA synthesis may reflect changes in the rate of transcription of a given set of genes but there is some evidence that increases in the rate of rRNA synthesis occur by the progressive activation of additional transcription units (McKnight and Miller, 1976).

Recent studies on plant rRNA genes has begun to elucidate the regulation of transcription by RNA polymerase I (Sollner-Webb and Tower, 1986; Sollner-Webb and Mougey, 1991; Hemleben and Zentgraf, 1994). Various regulatory elements suggested to be involved in RNA polymerase I transcription are located within the NTS region (Figure 1.4) which contains the transcription initiation and termination sites, the RNA polymerase I promoter, putative enhancer sequences and processing sites for the rRNA precursor, contained within an often complex organisation of 100-300 bp repeated elements (Hemleben and Zentgraf, 1994). As in *Xenopus* (Reeder *et al.*, 1983), the repeats are believed to serve as enhancers in RNA polymerase I transcription by factor trapping (Flavell *et al.*, 1986; Sardana

et al., 1993). When transcription components are limiting, the genes with more enhancers appear dominant over those with fewer enhancers. Moreover, a correlation has been found between the number of subrepeats and nucleolar activity in wheat (Flavell *et al.*, 1985).

1.6 GENES INVOLVED IN FIBRE DEVELOPMENT

The number of genes active in the developing cotton fibre is similar to that in other organs of the plant, and consists of both housekeeping and fibre-specific genes (Graves and Stewart, 1988a; John and Stewart, 1992). Although several fibre-specific genes have been isolated, mostly within the last two years, few have been characterised with respect to function and others show no sequence similarity to known proteins. In addition, several fibre mutants have been described, but their genetic basis, to date, is unknown.

1.6.1 *Arabidopsis* trichomes: an analogous system

Few cells elongate to the extent of cotton fibre cells in the absence of cell division. The only known analogous system in plants which has undergone extensive study is trichome development on the leaves of *Arabidopsis thaliana*. Like cotton fibres, *Arabidopsis* trichomes are single specialised epidermal cells which develop in a series of discrete developmental stages. The differentiation of trichomes of *Arabidopsis* has been utilised as a model for the study of plant cell differentiation and is well-characterised both at the genetic and molecular levels. The stages of growth are genetically defined by a bank of trichome mutants which affect a particular aspect of trichome development (Hülskamp *et al.*, 1994), generally either patterning or cell morphogenesis.

An interesting *Arabidopsis* mutant recently isolated shows altered patterns of cellulose deposition in particular organs so that deposition in the various cells of stems and roots is normal but the mutant is unable to synthesise and deposit cellulose in the secondary cell walls of trichomes (Potikha and Delmer, 1995). The mutation also affects callose deposition in

leaves, lending support to the suggestion that synthesis of cellulose and callose are related events. Potikha and Delmer (1995) suggest that the mutation occurs in a differentially-expressed gene which encodes a synthase subunit or in an activator associated with the tissue-specific expression. A similar tissue-specific activator, if present in cotton fibres, would be an obvious target in the genetic manipulation of fibre characteristics. Cytoskeletal organisation is not impaired and primary cell wall cellulose is deposited normally in mutant cells which suggests that cellulose synthesis and deposition are controlled independently in the two cell wall types, as is also suggested from studies of cotton fibres (Kohel *et al.*, 1974). A trichome-specific promoter has been identified and has potential in the expression of novel genes in the trichomes (Herman and Marks, 1989).

Interestingly, high levels of an XET protein (Section 1.3.3b) have been detected in *Arabidopsis* trichomes as well as in other growing regions of the plant (Xu *et al.*, 1995). The expression of the corresponding gene, *TCH4*, is regulated by expansion-promoting hormones and various environmental stimuli.

1.6.2 Characterisation of Genes Involved in Cotton Fibre Development

John and Crow (1992) were the first to report isolation of an abundant, fibre enhanced mRNA, E6. E6 transcripts were reported to be the most abundant fibre-specific transcripts in a fibre cell and were detected throughout fibre development, with a peak in expression at 15-22 DPA. Much lower levels of expression were observed in ovule, flower and leaf and the E6 promoter was shown to direct expression of a reporter gene in a tissue-specific and developmentally-regulated manner. The sequence of E6 did not show similarity to any known sequences and since it was present in fibre-producing rather than cellulose-producing organisms, it is postulated to have a unique structural or enzymatic role in the primary cell wall deposition in fibre (John and Crow, 1992). Two pistil-specific mRNAs recently isolated from *Brassica napus* have been found to share homology with E6 mRNA (Robert *et al.*, 1994). The

proteins encoded by the two *B. napus* genes are thought to have structural defence roles, protecting the surface of the stigma from pathogen infection.

Two further fibre-specific mRNAs of unknown function, B6 (John, 1995) and H6 (John and Keller, 1995), have been characterised. The B6 mRNA is present during both the primary and secondary wall stages of fibre growth and is also expressed at low levels in leaf tissue. The H6 mRNA appears to encode a proline-rich cell wall protein with features characteristic of AGPs (John and Keller, 1995; Section 1.4.2a), which is claimed to be the major proline-rich protein in cotton fibres. Based on its primary structure and expression pattern, the H6 protein is postulated to have a role in secondary cell wall deposition, perhaps in wall matrix assembly.

Several other fibre-specific mRNAs encode products with similarity to known proteins. Differential screening of a fibre cDNA library resulted in the isolation of an mRNA encoding a lipid transfer protein (LTP) (Ma *et al.*, 1995) which was present at low levels in leaves but was undetectable in roots and flowers. It has a signal peptide, suggesting that it is localised to the extracellular matrix and has been proposed to have a role in cuticle biosynthesis (Ma *et al.*, 1995). Two GTP-binding proteins show specific accumulation in cotton fibre cells and were isolated in a search for plant homologues of animal Rac proteins (Delmer *et al.*, 1995). Expression of one of the two genes peaks at about 15-18 DPA and a role in rearrangement of cytoskeletal elements during the transition between primary and secondary wall synthesis is envisaged.

Nine α-tubulin isotypes have been identified in cotton tissues and of these, two were found to accumulate specifically in fibres but were absent in other tissues (Dixon *et al.*, 1994). Similarly, two of the seven β-tubulin isotypes are specific to fibres and accumulate preferentially in secondary wall synthesis. In addition, the isotypes accumulated at different times during growth of the cotton fibre (Dixon *et al.*, 1994) and may correlate with the observed changes in cortical microtubule array during fibre development (Seagull, 1992).

1.7 Summary and Aims of the Project

Work on the cotton fibre is driven by economic, agricultural and scientific interests. The quality of cotton yarns and products is largely determined by the quality of the raw fibre and fibre properties, particularly length, strength and fineness, require further improvement to meet economic pressures and to facilitate more efficient spinning technology. Because these characters are primarily under genetic control, modification of fibre development through genetic manipulation could enhance fibre quality parameters. Prior to being able to utilise these techniques in cotton fibre improvement, an understanding of the messenger mRNAs and their protein products expressed in the differentiated cell is required. In addition to commercial applications, fibre-specific genes and their regulatory elements will be valuable tools to understand fibre differentiation and properties.

The primary aim of this project is to isolate and characterise cDNA clones corresponding to mRNAs which are present in fibre cells but absent in other differentiated cotton tissues. Overall protein synthesis, in the form of ribosome availability, has been implicated in cotton fibre development and the control of ribosomal gene activity was also investigated at the molecular level.

CHAPTER 2

Materials and Methods

2.1 MATERIALS

2.1.1 Samples of Plant Material

Sources of plant material utilised in this study are summarised in Table 2.1 and were seeds except where indicated.

2.1.2 Chemical Reagents

General laboratory chemicals were of analytical research grade and were purchased from a variety of manufacturers including Aldrich Chemical Co, BDH Chemicals Australia Pty Ltd, Oxoid Pty Ltd and Sigma Chemical Company. Specialist reagents used in this study are listed below:

"Amberlite" monobed resin	BDH Chemicals, Australia
Antifoam A emulsion	Sigma Chemical Co, St Louis, MO
β-mercaptoethanol	BDH Chemicals, Australia
Caesium chloride	BDH Chemicals, Australia
Chemicals for oligonucleotide synthesis	Beckman Instruments, Palo Alto, CA
DEPC	Sigma Chemical Co, St Louis, MO
Diphenylamine	BDH Chemicals, Australia
DMSO	BDH Chemicals, Australia
Guanidine thiocyanate	Fluka Biochemika
PEG 8000	Sigma Chemical Co, St Louis, MO
PVP	Fluka Biochemika
ReadySolv™HP	Beckman Instruments, Palo Alto, CA
Sodium borate decahydrate	Sigma Chemical Co, St Louis, MO
Spermidine	Sigma Chemical Co, St Louis, MO
TEMED	BDH Chemicals, Australia
Urea	BDH Chemicals, Australia

2.1.3 Stains and Dyes

Bromophenol blue	Sigma Chemical Co, St Louis, MO
Coomassie blue	Gibco BRL, Gaithersburg, MD
Ethidium bromide	Sigma Chemical Co, St Louis, MO
Methyl green	George T Gurr Ltd, London

Table 2.1: Plant material utilised in this study

Species	Source	Location
Cotton varieties		
G. hirsutum L., cv. Deltapine 90	Cotton Seed Distributors, Narrabri, NSW	Chapter 6
G. hirsutum L., cv. Siokra 1-2	Cotton Seed Distributors, Narrabri, NSW	Chapters 3, 4, 5, 6
G. hirsutum L., cv. Siokra 1-4	Cotton Seed Distributors, Narrabri, NSW	Chapters 5, 6
G. hirsutum L., cv. Siokra L-22	Cotton Seed Distributors, Narrabri, NSW	Chapters 5, 6
G. hirsutum L., cv. CS7S	Cotton Seed Distributors, Narrabri, NSW	Chapters 5, 6
G. hirsutum L., cv. Strumica 5086	Dr Greg Constable (Cotton Research and Development Corporation, Narrabri, NSW)	Chapters 5, 6
G. hirsutum L., cv. Paymaster 909	Dr Greg Constable (Cotton Research and Development Corporation, Narrabri, NSW)	Chapters 5, 6
G. barbadense L., cv. Pima S-6	Richard MacCullum (Lake Tandau, Menindee, NSW)	Chapters 5, 6
G. raimondii Ulbr.	Genomic DNA from two plants; Dr Curt Brubaker (CSIRO Division of Plant Industry, Canberra, ACT)	Chapter 5
G. herbaceum L.	Leaves from three plants of accession LL1776; Dr Curt Brubaker (CSIRO Division of Plant Industry, Canberra, ACT)	Chapter 5
G. sturtianum J. H. Willis	Nindethana Seed Service, WA	Chapter 5
G. robinsonii F. Muell.	Nindethana Seed Service, WA	Chapter 5
Other plants		
Arabidopsis thaliana ecotype Columbia	Genomic DNA; Dr Daniel Schachtman, Department of Botany, The University of Adelaide	Chapter 5
Pea, *Pisum sativum*	Leaves; Dr Joe Wiskich, Department of Botany, The University of Adelaide	Chapter 5
Carrot, *Daucus carota*	Grown from existing seeds in the department	Chapter 5
Tomato, *Lycopersicon esculentum*	Grown from existing seeds in the department	Chapter 5
Tobacco, *Nicotiana tabacum*	Leaves; Sabba Spiniello, Department of Genetics, The University of Adelaide	Chapter 5
Silverbeet, *Beta vulgaris* var. silverbeet	Grown from existing seeds in the department	Chapter 5

PyroninY	ICN Pharmaceuticals Inc, Costa Mesa, CA
Xylene cyanole FF	Bio-Rad Laboratories

2.1.4 Enzymes

Restriction endonucleases (and 10 x restriction buffers) were obtained from Promega and Boehringer Mannheim. The remaining enzymes used in this study were obtained from the following manufacturers:

Calf intestinal phosphatase	Boehringer Mannheim, Germany
DNaseI (grade 1)	Boehringer Mannheim, Germany
Lysozyme	Sigma Chemical Co, St Louis, MO
M-MLV reverse transcriptase	Gibco BRL, Gaithersburg, MD
Proteinase K	Boehringer Mannheim, Germany
RNaseA	Sigma Chemical Co, St Louis, MO
RNasin®	Promega, Madison, WI
T_4 DNA ligase	New England Biolabs, Beverly, MA
T_4 polynucleotide kinase	Boehringer Mannheim, Germany
Taq DNA polymerase	Bresatec, Adelaide, Australia

2.1.5 Radioactive Isotopes

[α-^{32}P]dATP (3000 Ci/mmole; ethanolic)	Bresatec, Adelaide, Australia
[α-^{32}P]dCTP (3000 Ci/mmole; aqueous)	Bresatec, Adelaide, Australia
[α-^{32}P]dATP (3000 Ci/mmole; Redivue)	Radiochemical Centre, Amersham, UK
[α-^{33}P]dATP (1000-3000 Ci/mmole; Redivue)	Radiochemical Centre, Amersham, UK
[α-^{35}S]dATP (1000 Ci/mmole; Redivue)	Radiochemical Centre, Amersham, UK
[γ-^{32}P]dATP (3000 Ci/mmole; Redivue)	Radiochemical Centre, Amersham, UK
[L-^{35}S]methionine (1000 Ci/mM)	Radiochemical Centre, Amersham, UK

2.1.6 Nucleic Acids

A number of plasmid clones and oligonucleotides were used in this study and are presented in Table 2.2 and Table 2.3 respectively. Note that due to ambiguity in nomenclature of sense and antisense DNA strands, the coding strand of a gene will be referred to as such throughout this thesis whilst the opposite strand will be referred to as the mRNA-like strand. Molecular weight markers and other nucleic acids were purchased from the following suppliers:

Calf thymus DNA	Sigma Chemical Co, St Louis, MO
DNA ladder, 100 bp	Promega, Madison, WI
Lambda DNA restricted with *Hin*dIII	Progen Industries Ltd, Australia
oligo$(dA)_{12-18}$	Pharmacia, Uppsala, Sweden
oligo$(dT)_{12-18}$	Pharmacia, Uppsala, Sweden
pUC19 DNA restricted with *Hpa*II	Biotech International Ltd, Australia
RNA size markers	Promega, Madison, WI
Salmon sperm DNA	Sigma Chemical Co, St Louis, MO
Unlabelled deoxyribonucleotides	Boehringer-Mannheim, Germany

2.1.7 Cloning Vectors

The plasmid and bacteriophage vectors used in this study are presented in Table 2.4

2.1.8 Bacterial Strains

The bacterial strains used in this study are presented in Table 2.5

2.1.9 Antibiotics and Indicators

Ampicillin	Boehringer-Mannheim, Germany
Chloramphenicol	Progen Industries Ltd, Australia
IPTG	Boehringer-Mannheim, Germany
Kanamycin	Sigma Chemical Co, St Louis, MO
Tetracycline	Sigma Chemical Co, St Louis, MO
X-gal	Progen Industries Ltd, Australia

Table 2.2: Plasmid clones

Plasmid	Insert	Vector	Location	Reference
pCU18	A 10 kb fragment of cucumber *Cucumis sativus* rDNA, including 25S, 18S genes and transcribed spacers	pACYC184	Chapter 6	Kavanagh and Timmis, 1986
pF15	A 130 bp *Bam*HI/*Pma*CI fragment of rye grass *Secale cereale* 5S rDNA	pBluescript®KS(±)	Chapter 4	Lawrence and Appels, 1986
Xho3	A 3.1 kb fragment of spinach *Spinacia oleraceae* chloroplast rDNA, including 23S gene and intergenic spacer	pACYC177	Chapter 6	Palmer and Thompson, 1981; Crouse *et al.*, 1978
910215-34	Complete rDNA repeat unit of grapevine *Vitis champini*	pBluescript®SK(+)	Chapter 6	

Table 2.3a: Oligonucleotide primers to vector sequences

Primer	Sequence (5'→3')	Source and Use	Location
Plasmid primers			
T3	ATT AAC CCT CAC TAA AG	PCR primer for determination of insert orientation in pBluescript® clones	Chapter 4, 5
T7	AAT ACG ACT CAC TAT AG	PCR primer for determination of insert orientation in pBluescript®, pGEM®-T and λGEM-11™ clones	Chapter 4, 5
RSP	AAC AGC TAT GAC CAT G	New England Biolabs, Beverly, MA; Sequencing of pBluescript® inserts and PCR primer for determination of insert orientation in pGEM®-T clones	Chapters 3, 4, 5, 6
(-40)	GTT TTC CCA GTC ACG AC	Bresatec, Adelaide, Australia; Sequencing of pBluescript® inserts	Chapters 3, 4, 5, 6
Phage primers			
ON-L	AGG TCG CCG CCC	Promega, Madison, WI; Mapping of λGEM-11™ genomic clones by partial restriction	Chapter 3, 4, 5
ON-R	GGG CGG CGA CCT	Promega, Madison, WI; Mapping of λGEM-11™ genomic clones by partial restriction	Chapter 3, 4, 5

Table 2.3b: Oligonucleotide primers designed from insert sequences

Primer	Sequence (5'→3')	Use	Location
Primers for 5'-RACE			
P3.1	GATTAGAGGCTACTCCCCTCATCACC	First strand cDNA synthesis for pFS3	Chapter 4
P3.2	GGGAATTCACCAAGAGAGCTATTAAAGAGACC	5'-RACE (PCR) for pFS3	Chapter 4
P21.1	GGCTTCCACTAGCTGGTGAACTTTCC	First strand cDNA synthesis for pFS6 group clones	Chapter 5
P21.2	GGGAATTCCAAAACTTCACTTGACGCTGTTGC	5'-RACE (PCR) for pFS6 group clones	Chapter 5
P17.1	CATTCGAATGAGTACAGCCATGAGC	First strand cDNA synthesis for pFS17	Chapter 4
P17.2	GGGAATTCGCAACTGGTCTCTCTTCAAACATG	5'-RACE (PCR) for pFS17	Chapter 4
P17R2	CACCAGTTTATACACC	Further sequencing of pFS17 5'-RACE clones in reverse direction, generation of a PCR product for subcloning	Chapter 4
P17R3	AGAAGCCGAGCAAAGCACGG	Further sequencing of pFS17 5'-RACE clones in reverse direction, generation of a PCR product for subcloning	Chapter 4
P17F2	CCTAGGTGGTTTTGGC	Further sequencing of pFS17 5'-RACE clones in forward direction	Chapter 4
Primers for LTP genes (pX2.2)			
ltpseq1	CCATGCACAACACCGCCACACATG	Sequencing into promoter; binds mRNA-like DNA strand	Chapter 5
ltpseq2	TTTTATGAGCCCTTTCCACC	Extension of sequence from ltpseq1 further into promoter; binds mRNA-like DNA strand	Chapter 5
P21F2	TGTCTCCAAAAGAGTCGTCC	Extension of sequence from (-40) primer to 3' end of gene; binds mRNA-like DNA strand	Chapter 5
P21R2	AATAATGTGGTGGAAAGGGC	Extension of sequence from RSP to 5' end of gene; binds coding DNA strand	Chapter 5
ltpseq-1	CGTGTTTCTCAGTTCGTTCC	Extension of sequence from P21R2 into 5' coding region; binds coding DNA strand	Chapter 5
P21-1	GCAAATGCATCAAAAGTGCG	Sequencing into 3' end of gene; binds coding DNA strand	Chapter 5
P21-2	CACTGACTGCAGCAGGTTCG	Extension of sequence from P21-1 into 3'-UTR; binds coding DNA strand	Chapter 5
Sundries			
EFP1	GAGATTATTTGAGAGAG	PCR amplification of 3' end of pGhEF-1α	Chapter 3

Table 2.4: Plasmid and bacteriophage vectors

Vector	Source	Use	Reference
Plasmids			
pGEM®-T	Promega, Madison, WI	Cloning of PCR products	Robles and Doers, 1994
Phagemids			
pBluescript®SK(-)	Stratagene GmbH, Heidelburg, Germany (excision from λZAP®II)	General cloning and subcloning vector	Short *et al.*, 1988
Phage			
λGEM-11™	λGEM-11™ *Bam*HI arms cloning system; Promega, Madison, WI	Construction of a partial cotton genomic library	Frischauf *et al.*, 1983; Karn *et al.*, 1984
λZAP®II	λZAP®II digested with *Eco*RI and dephosphorylaetd; Stratagene GmbH, Heidelburg, Germany	Construction of a cotton fibre cDNA library; cloning of pFS17 5'-RACE product	Short *et al.*, 1988
ExAssist™ helper filamentous phage	Stratagene GmbH, Heidelburg, Germany	Rescue of phagemids from λZAP®II vector	Hay and Short, 1992

Table 2.5: Bacterial strains (*Escherichia coli*)

E. coli Strain	Source	Use	Genotype and Reference
DH-5αF'	Laboratory stock	Host for cloning into pGEM®-T and general cloning	*recA*1, *endA*1, *gyrA*96, *thi*-1, *hsdR*17 (r_K^-, m_K^+), *supE*44, *relA*1, *deoR*, Δ(*lacZYA-argF*)U169 [F', Φ80d*lacZ*ΔM15]; Hanahan, 1983
KW251	Promega, Madison, WI	Host for propagation of λGEM-11™ cotton genomic library	*supE*44, *galK*2, *galT*22, *metB*1, *hsdR*2, *mcrB*1, *mcr*A, [*arg*A81: Tn10], *recD*1014
PLK-F'	Stratagene GmbH, Heidelburg, Germany	Host for cloning of ribosomal rRNA repeat unit from cotton	*recA1*, *lac*$^-$, *mcrA*$^-$, *mcrB*$^-$, *hsd*, (*rk*$^-$, *mk*$^+$). *gal*$^-$, *supE* [F', *proAB*, *lacI*^{q}ZΔ*M15*, Tn*10* (*tet*^{r})]; Yanisch-Perron *et al.*, 1985
SOLR™	Stratagene GmbH, Heidelburg, Germany	Host for excision of phagemid from λZAP®II vector	e14$^-$(*mcr*A), Δ(*mcrCB-hsdSMR-mrr*)171, *sbcC*, *recB*, *recJ*, *umuC*::Tn*5*(kan^r), *uvrC*, *lac*, *gyrA96*, *relA1*, *thi-1*, *endA1*, λ^R, [F' *proAB*, *lacI*^{q}Z*ΔM15*]Su$^-$; Hay and Short, 1992
XL1-Blue	Stratagene GmbH, Heidelburg, Germany	Host for propagation of λZAP®II fibre cDNA library	*recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*, *supE44*, *relA1*, *lac*, [F' *proAB*, *lacI*^{q}Z Δ*M15*, Tn*10*(tet^r)]; Bullock *et al.*, 1987

2.1.10 Kits and Miscellaneous Materials

Kits and other materials used in this study were obtained from the following manufacturers:

5'-AmpliFINDER™ RACE kit	CLONTECH Laboratories Inc, Palo Alto, CA
Bio-Gel P60 (50-100 mesh)	Bio-Rad Laboratories, Richmond, CA
Bio-Gel P60 (100-200 mesh)	Bio-Rad Laboratories, Richmond, CA
cDNA Synthesis Kit	Pharmacia, Uppsala, Sweden
DNA UltraFast Cleavage and Deprotection Kit	Beckman Instruments, Palo Alto, CA
Exassist™/SOLR™ System	Stratagene, La Jolla, CA
fmol™ DNA Sequencing System	Promega, Madison, WI
GIGAprime DNA Labelling Kit	Bresatec, Adelaide, Australia
Glass beads, 150-212 µm	Sigma Chemical Co, St Louis, MO
Hybond™-N+ nylon membrane	Amersham, UK
Hybond™mAP messenger affinity paper	Amersham, UK
LambdaMap™ System	Promega, Madison, WI
mRNA Purification Kit	Pharmacia, Uppsala, Sweden
Nitrocellulose	Schleicher and Schuell GmbH, Dassel, Germany
Packagene® Lambda DNA Packaging System	Promega, Madison, WI
pGEM®-T Vector System	Promega, Madison, WI
PolyATtract® mRNA Isolation System	Promega, Madison, WI
Sepharose CL-6B	Pharmacia, Uppsala, Sweden
Sequenase Version 2.0 DNA Sequencing Kit	United States Biochemical, Cleveland, OH
TNT™ T3 Coupled Reticulocyte Lysate System	Promega, Madison, WI
Ultrafree®-MC 0.45 µm filter unit	Millipore, Bedford, MA
X-ray film	Fuji RX Medical X-ray Film, Fuji Photo Film Co, Ltd

2.2 SOLUTIONS, BUFFERS AND MEDIA

2.2.1 Solutions and Buffers

All solutions and buffers were prepared using millipore water and autoclaved where possible, otherwise were sterilised by passage through a 0.2 µm filter. Solutions for RNA work were treated with 0.1% (v/v) DEPC as per Sambrook *et al.* (1989) prior to autoclaving. All restriction endonucleases and most other enzymes were provided with incubation buffers. Solutions and all other buffers routinely used in this study were as follows:

Acid mix (slotblots)	1 volume 1M HCl, 1 volume 1M Tris-HCl pH 8.0, 2 volumes 3M NaCl
Acrylamide stock	38% (w/v) acrylamide and 2% (w/v) bis-acrylamide, deionised with monobed resin
Denaturing solution	1.5M NaCl, 0.5M NaOH
1 x Denhardt's	0.02% (w/v) Ficoll®, 0.02% (w/v) PVP, 0.02% (w/v) gelatin
Diphenylamine reagent	1.5% (w/v) diphenylamine in glacial acetic acid, with the addition of 1.5 ml H_2SO_4
DNA extraction buffer	2 x SSC, 50mM EDTA pH 8.0, 2% (w/v) sarcosyl
6 x DNA loading buffer	0.25% (w/v) bromophenol blue, 40% (w/v) sucrose
Formamide RNA loading buffer	50% (v/v) formamide, 17.5% formaldehyde, 10% (v/v) 10 x MOPS buffer, 50 mg (w/v) ethidium bromide
Hybridisation buffer	5 x SSPE, 5 x Denhardt's, 0.5% (w/v) SDS, 100 mg/ml denatured, sonicated salmon sperm DNA
Lysing solution	10mM glucose, 25mM Tris-HCl pH 8.0, 10mM EDTA, 4 mg/ml lysozyme
Methyl green-pyroninY	0.5 %(w/v) methyl green, 0.1M Na-acetate pH 4.4, repeated chloroform extraction to remove residual methyl violet, followed by addition of 0.2% (w/v) pyroninY
10 x MOPS buffer	200mM MOPS, 50mM anhydrous Na-acetate, 10mM EDTA, adjusted to pH 7.0

Neutralising solution	1.5M NaCl, 1M Tris-HCl pH 7.2, 1mM EDTA
Neutral mix (slotblots)	1 volume 1M NaOH, 1 volume 1 x TE, 4 volumes acid mix
Phenol/Chloroform	50% (v/v) phenol, 48% (v/v) chloroform, 2% (v/v) isoamylalcohol, buffered with an equal volume of Tris-HCl pH 8.0, 0.2% (v/v) β-mercaptoethanol
10 x polynucleotide kinase buffer	0.5M Tris-HCL pH 7.6, 0.1M $MgCl_2$, 50mM DTT, 1mM spermidine, 1mM EDTA
RNA homogenisation buffer	5M guanidine isothiocyanate, 0.2M Tris-acetate pH 8.5, 0.2% (w/v) sarcosyl, with 0.7% (v/v) β-mercaptoethanol and 1% (w/v) PVP freshly added
5 x RNA tracking dye	8% (w/v) Ficoll, 0.02% (w/v) bromophenol blue, 0.04% (w/v) xylene cyanole FF
SM buffer	100mM NaCl, 10mM $MgSO_4$, 50mM Tris-HCl pH 7.5, 2% (w/v) gelatin
1 x SSC	0.15M NaCl, 0.015M Na_3-citrate, adjusted to pH 7.2 with NaOH
1 x SSPE	0.18M NaCl, 10mM NaH_2PO_4, 1mM EDTA pH 7.2
10 x STE	100mM Tris-HCl pH 8.0, 1M NaCl, 10mM EDTA
Stop buffer	10 mM Tris-HCl pH 7.5, 1mM EDTA pH 8.0, 0.2% (w/v) SDS
1 x TAE	40mM Tris-HCl, 20mM Na-acetate, 2mM EDTA, adjusted to pH 7.8 with glacial acetic acid
1 x TBE	0.9M Tris-borate, 0.002M EDTA pH 8.0
1 x TE	10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0

2.2.2 Media

All liquid and solid media were prepared using millipore water and sterilised immediately by autoclaving. The compositions of various media were as follows:

Luria broth (L-broth)	1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, adjusted to pH 7.2 with 1M NaOH

Luria agar (L-agar)	Luria broth with the addition of 1% (w/v) bacteriological agar No. 1
NZY agar	1% (w/v) NZ amine (casein hydrolysate), 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 1% (w/v) bacteriological agar No. 1
NZY top agar/agarose	1% (w/v) NZ amine (casein hydrolysate), 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 10mM $MgSO_4$, 0.7% (w/v) bacteriological agar No. 1 or agarose
TB agar	1% (w/v) tryptone, 0.5% (w/v) NaCl, 1% (w/v) bacteriological agar No. 1
TB top agar/agarose	1% (w/v) tryptone, 0.5% (w/v) NaCl, 0.01M $MgSO_4$, 0.7% (w/v) bacteriological agar No. 1 or agarose
2 x YT broth	1.6% (w/v) tryptone, 1% (w/v) NaCl, 1% (w/v) yeast extract, adjusted to pH 7.0 with 1M NaOH

2.3 METHODS

2.3.1 Plant Growth Conditions and Fibre Harvest

Cotton seeds were sown in soil at a density of 2-4 plants per 8 L pot and propagated in growth cabinets under cycling conditions (day/night temperatures of 30°C/25°C, light/dark cycle of 16/8 hr). The light source was fluorescent and incandescent lamps. All other plants were grown in the greenhouse and seedlings were fertilised periodically with Osmocote. Seeds of Australian native *Gossypium* species, *G. sturtianum* and *G. robinsonii*, were soaked in dH_2O overnight before dissection of the embryos from the seed coat. The embryos were germinated on Whatman® 3MM paper in sterile conditions and in the dark for 2 days, then in the light for 4 days before planting in soil.

Cotton flowers were tagged on the day of anthesis. Bolls were harvested at various days postanthesis (DPA) and stored on ice before opening with a sharp knife. Fibres were immediately removed from the seeds under cold dH_2O using fine forceps, blotted on paper towel and frozen in liquid nitrogen. Care was taken to avoid removal of the fuzz hairs or the

seed coat. The yield from each boll was weighed before storage at -70°C. Leaf and whole flowers were also collected from mature plants and snap-frozen in liquid nitrogen before storage at -70°C until further use. Root and hypocotyl (seedling) tissue was harvested when they had reached a height of approximately 5 cm.

2.3.2 General Techniques

2.3.2a Restriction Enzyme Digests

All reactions involving restriction endonucleases were performed under conditions as recommended by the manufacturers (Boehringer Mannheim or Promega). In general, five units of enzyme were added per µg of DNA to a maximum enzyme volume of 20% of the total reaction volume and the reaction was allowed to proceed for 2-3 hr. Spermidine to a final concentration of 3mM and RNaseA to a final concentration of 0.5 µg/µl were added to all restrictions. Samples of plant genomic DNA were restricted with 10 units of enzyme per µg of DNA for at least 16 hr.

2.3.2b Electrophoretic Separation of DNA

DNA was size fractionated by electrophoresis on 0.7% (w/v) agarose gels in submarine gel tanks with 1 x TAE as the running buffer. Amplified DNA samples were analysed by electrophoresis on 1.2-1.5% (w/v) agarose gels. Samples were loaded with 1 x loading buffer and electrophoresed at 60-80 V until the bromophenol blue marker dye had migrated a sufficient distance to allow adequate separation of the DNA fragments. Following electrophoresis, the DNA was stained in 5 µg/ml of ethidium bromide and viewed under UV light. Genomic DNA gels to be Southern blotted were electrophoresed at 30 V for at least 16 hr. Where applicable, sizes of DNA fragments were estimated using the Fragrap computer program (Duggleby *et al.*, 1981).

2.3.2c Recovery of DNA from Agarose Gels

DNA was eluted from agarose gels using the freeze-squeeze method adapted from Thuring *et al.* (1975). The frozen gel slice was squeezed through a 26G syringe needle into the upper compartment of an Ultrafree®-MC 0.45 µm filter unit (Millipore) and washed through with 200 µl of 1 x TAE buffer. The DNA was allowed to elute from the gel at room temperature for 15 min. Centrifugation for 2-3 min at 7000 x g yielded a solution of DNA in the lower tube compartment and the DNA was recovered by precipitation with ethanol.

2.3.2d Precipitation of DNA

Method 1 : Precipitation by Ethanol

DNA was precipitated by the addition of one tenth of the volume of 3M Na-acetate pH 5.2 followed by 2-2.5 volumes of cold redistilled ethanol. The DNA was left to precipitate at -20°C for a period of 2-16 hr depending on its molecular weight. The DNA was recovered by centrifugation at 12000 x g for at least 15 min at 4°C. The pellet was washed with cold 70% (v/v) ethanol, dried under vacuum and resuspended in a minimal volume of 1 x TE.

Method 2 : Precipitation by Isopropanol

In some cases, DNA was precipitated by addition of one tenth of the volume of 3M Na-acetate pH 5.2 followed by an equal volume of cold isopropanol. The DNA was left to precipitate at room temperature for 5 min before microcentrifugation at 12000 x g for at least 15 min. The resulting DNA pellet was washed with cold 70% (v/v) ethanol, dried under vacuum and resuspended in a minimal volume of 1 x TE.

2.3.2e Purification of DNA Using Phenol/Chloroform Extraction

DNA was purified from proteins by the addition of an equal volume of phenol/chloroform and vigorous mixing. The emulsion was centrifuged at 12000 x g for 2 min and the aqueous layer transferred to a new tube before the extraction was repeated.

Contaminating phenol was removed by the addition of an equal volume of chloroform, mixing and centrifugation as above, and the DNA solution transferred to a new tube.

2.3.2f Purification of DNA Using a CsCl/EtBr Gradient

DNA from large-scale plasmid and genomic DNA preparations was purified by density equilibrium centrifugation in a CsCl/EtBr gradient. Genomic and plasmid DNA solutions to be purified were made up to 5 ml with 1 x TE and 5 g of CsCl was added and dissolved. Following addition of 200 µl of 10 mg/ml EtBr, gradients were centrifuged at 40000 rpm for 40 hr at 20°C (using the Ti70.1 rotor of a Beckman ultracentrifuge). The DNA was recovered using methods described by Sambrook *et al.*(1989) and EtBr removed by butanol extraction. The DNA was precipitated by addition of ethanol.

2.3.2g Determination of DNA and RNA Concentration

The concentration of DNA and RNA in solution was determined by UV spectrophotometry at a wavelength of 260 ηm, or in the case of freeze-squeezed restriction fragments, by gel electrophoresis of a sample in conjunction with *Hin*dIII-digested Lambda DNA markers.

2.3.3 Preparation of DNA Samples

2.3.3a Isolation of Genomic DNA from Plants

Total genomic DNA from the young leaves of cotton and other plants was isolated by the method of Dellaporta *et al.* (1985). *G. raimondii* genomic DNA was purified by phenol/chloroform extraction and precipitation with isopropanol.

2.3.3b Isolation of Genomic DNA for Construction of a Genomic Library

The protocol followed for extraction of DNA from Siokra 1-2 leaf tissue for construction of a genomic library was a modification of the method used by Scott and Possingham (1980). Approximately 5 g of leaf material was ground to a fine powder in liquid

nitrogen and thawed into 25 ml of DNA extraction buffer. The suspension was incubated with 100 µg RNaseA and 100 µg Proteinase K for 1 hr at 37°C. The supernatant from a 10 min, 15000 x g centrifugation was filtered through 1 layer of miracloth and then subjected to centrifugation at 60000 rpm for 3 hr at 5°C (using the Ti70 rotor of a Beckman ultracentrifuge). The resulting pellet was dissolved in 1 ml of 1 x TE and the DNA purified using a CsCl/EtBr gradient.

2.3.3c Small-scale Plasmid Preparation

Method 1 : Alkaline Lysis Plasmid Minipreps

Plasmid DNA for general applications was prepared using the scaled-up version of the alkaline lysis method (for 10 ml of overnight bacterial culture) as per Sambrook *et al.*(1989).

Method 2 : Lithium Chloride Method

Phagemid DNA from rescue of recombinant "cold" plaques was isolated using the 20-minute lithium chloride protocol as described by He *et al.* (1990).

Method 3 : Boiling Lysis

Plasmid DNA for sequencing was prepared using the boiling lysis method of Holmes and Quigley (1981), scaled down for 1.5 ml of a 10 ml overnight bacterial culture.

2.3.3d Large-scale Plasmid Preparation

The method used for large-scale preparation of plasmid DNA was adapted from the alkaline lysis method of Sambrook *et al.* (1989). The medium used for all bacterial cultures was L-broth containing 100 µg/ml of ampicillin, and all growth stages were performed at 37°C on a shaking platform.

A 1 ml aliquot from a 10 ml overnight culture of bacteria was used to inoculate 50 ml of medium which was grown until OD_{600} = 0.6. A sample of 25 ml of this culture was then added to 500 ml of medium and grown until OD_{600} = 0.9. Chloramphenicol (75 mg) was added to

amplify the plasmid and the cells incubated overnight. The bacterial pellet from centrifugation for 10 min at 7000 x g was resuspended in 10 ml of 50mM glucose, 10mM EDTA, 25mM Tris-HCl pH 8.0, 20 mg of lysozyme added and the mixture incubated for 30 min on ice. To complete cell lysis, 20 ml of 0.2M NaOH, 1% SDS were added to the cells and the mixture allowed to stand 5 min on ice. The lysed cells were incubated on ice for a further 60 min after addition of 15 ml of 3M Na-acetate pH 4.8. Centrifugation at 27000 x g for 20 min yielded a supernatant from which the plasmid DNA was recovered by ethanol precipitation. Plasmid DNA was resuspended in 5 ml of 1 x TE buffer and purified using a CsCl/EtBr gradient.

2.3.3e Preparation of Lambda DNA From Plate Lysates

Recombinant λGEM-11™ DNA was prepared by a modification of the method reported in Sambrook *et al.* (1989). Lambda phage were plated to confluence on TB agarose plates. The plates were each washed with 4 ml of SM buffer at room temperature for 2 hr with gentle shaking. The solution from four plates was combined into a 15 ml Corex® tube and each plate rinsed with an extra 1 ml of SM buffer before centrifugation at 8000 x g for 10 min to pellet bacterial debris. The supernatant was transferred to a fresh tube, DNaseI and RNaseA were added to final concentrations of 0.1 µg/ml and 1 µg/ml respectively, and the mixture incubated at 37°C for 60 min. An equal volume of cold 2M NaCl, 20% (w/v) PEG 8000 was added and the mixture incubated on ice for 60 min. Phage particles were collected by centrifugation at 11000 x g for 20 min and the pellet drained to ensure the complete removal of PEG. The pellet was resuspended in 0.5 ml of SM buffer by vigorous vortexing and insoluble debris removed by centrifugation at 8000 x g for 2 min. Following addition of 5 µl 0.5M EDTA and 5 µl 10% (w/v) SDS to the supernatant, it was incubated at 68°C for 20 min to lyse the phage. The solution was extracted twice with water-saturated phenol, once with phenol/chloroform and once with chloroform. Phage DNA was recovered by the addition of isopropanol, resuspended in 100 µl 0.3M Na-acetate and reprecipitated with ethanol, and the resultant DNA pellet was

redissolved in 50 µl 1 x TE buffer.

2.3.4 Cloning Strategies

2.3.4a Preparation of Insert DNA and Plasmid Vector

DNA fragments for ligation were generated in one of two ways. DNA was restricted to completion with the appropriate restriction enzyme and then recovered from agarose gels using the freeze-squeeze method or purified by phenol/chloroform extraction and precipitation with ethanol.

Vector DNA for use in ligations was digested with the appropriate restriction enzyme, purified by extraction with phenol/chloroform and precipitated with ethanol. Completely digested plasmid DNA, analysed on an agarose gel, was treated with calf intestinal phosphatase (CIP) essentially as per Sambrook *et al.* (1989) in order to prevent self-ligation. The 50 µl reaction was incubated at 37°C for 30 min, after which 10 µl 10 x STE, 5 µl 10% (v/v) SDS and 40 µl dH_2O were added and the mixture incubated at 68°C for 15 min to deactivate the enzyme. A phenol/chloroform extraction was carried out and the DNA precipitated by the addition of ethanol. The final concentration of the vector was 50 ηg/µl.

2.3.4b Cloning of Products Generated by PCR

5'-RACE PCR products were purified by centrifugation through a Sepharose CL-6B chromatography column (Murphy and Kavanagh, 1988) prior to restriction with *Eco*RI, purification by phenol/chloroform extraction and precipitation with ethanol. pFS3 and pFS6 5'-RACE PCR products were cloned into pBluescript®SK(-), whilst the pFS17 5'-RACE PCR product was purified from an agarose gel prior to the above restriction and purification steps and ligation into *Eco*RI-digested and dephosphorylated λZAP®II vector (Stratagene). All PCR products derived from the pFS17 5'-RACE clones were recovered from agarose gels before cloning.

2.3.4c Ligations

Ligations were performed using 50-100 ηg of vector DNA and an equimolar amount of insert DNA, catalysed by 1 Weiss unit of T_4 DNA ligase in a 10 µl reaction. All ligations were placed in a 7 L insulated waterbath at 16°C and left overnight in a 4°C coldroom. Products which resulted from amplification of pFS17 5'-RACE clones were ligated into pGEM®-T (Promega) following manufacturer's instructions and utilising New England Biolabs T_4 DNA ligase.

2.3.4d Preparation and Transformation of Competent *E. coli*

Preparation of DH5αF' and PLK-F' competent cells for transformation was performed using the calcium chloride method as per Sambrook *et al.* (1989). Transformation of competent *E. coli* cells with ligation mixtures was achieved using heat shock at 42°C as per Sambrook *et al.* (1989). The cells were pelleted at 7000 x g for 30 sec, resuspended in 400 µl L-broth and plated onto L-agar plates supplemented with 100 µg/ml ampicillin. Colonies containing recombinant plasmids were selected by plating onto L-agar medium containing 100 µg/ml ampicillin with an overlayer of L-agar plus 100 µg/ml ampicillin, 200 µg/ml X-gal and 160 µg/ml IPTG.

2.3.5 RNA Manipulation

2.3.5a Total RNA Preparation

RNA for the construction of the cDNA library from total RNA (Chapter 3) was isolated from cotton fibres and other tissues using a method modified from that of John (1992). Cotton tissue (1-2 g) was powdered in a mortar in liquid nitrogen, thawed into RNA homogenisation buffer and homogenised using a polytron on full speed for 1.5 min. Following centrifugation at 10000 x g for 5 min at room temperature, the supernatant was filtered through 2 layers of miracloth which had been pre-wet with homogenisation buffer. Samples of 3 ml were

transferred to polyallomer ultracentrifuge tubes and underlaid with 2 ml of 5.7M CsCl, 100mM EDTA pH 7.0. The tubes were centrifuged at 40000 rpm for 18 hr at 20°C (using the SW55Ti rotor of a Beckman ultracentrifuge). The supernatant was aspirated from the RNA pellet which was then dissolved in 0.5 ml dH_2O to which 100 U RNAsin® and DTT to a final concentration of 2mM had been added. Following removal of a sample for UV spectrophotometry, the RNA was precipitated in 10 µg aliquots and stored at -70°C. The integrity of this RNA was verified by *in vitro* translation, using the TNT™ T3 Coupled Reticulocyte Lysate System (Promega) and [L-^{35}S]methionine (Amersham). Products from the reaction were analysed by SDS-polyacrylamide gel electrophoresis on a Mighty Small II (Hoefer Scientific Instruments) apparatus using established protocols (Laemmli, 1970; Sambrook *et al.*, 1989).

RNA for the subsequent cDNA library (Chapter 4) was prepared from all cotton tissues using the method of Wan and Wilkins (1994), except that amounts of all reagents were doubled to accommodate 1 g of tissue and a polytron was used in place of a glass tissue grinder. RNA was stored in aliquots under 80% (v/v) ethanol at -70°C.

2.3.5b Poly(A)$^+$ RNA Preparation

Attempts to purify polyadenylated (poly(A)$^+$) RNA from 10-24 DPA fibre utilised Hybond™mAP messenger affinity paper (Amersham) and the PolyATtract® mRNA Isolation System (Promega) and were performed according to the manufacturer's instructions.

Poly(A)$^+$ RNA from leaf and 13 DPA fibre was affinity purified on oligo(dT)-cellulose columns (mRNA Purification Kit, Pharmacia) according to the protocol from the manufacturer.

2.3.5c Electrophoretic Separation of RNA

Aliquots of 10 µg of total RNA were size fractionated by electrophoresis on denaturing formaldehyde agarose gels. RNA samples under 80% ethanol were recovered by centrifugation

at 12000 x g for 20 min at 4°C, dried *in vacuo* and dissolved in 20 µl of formamide RNA loading buffer. Following denaturation at 65°C for 5 min, samples were transferred to ice and loaded with 1 x RNA tracking dye into a 1.4% (w/v) agarose, 15% (v/v) formaldehyde, 1 x MOPS gel. The gel was run at 50 mA with 1 x MOPS as the running buffer and viewed under UV light.

2.3.6 Construction and Screening of Cotton Fibre cDNA Libraries

Double-stranded cDNA was synthesised from total RNA (Chapter 3) and poly(A)$^+$ RNA (Chapter 4) using a Pharmacia cDNA Synthesis Kit, according to manufacturer's instructions except that cDNA was passed through two spun columns to ensure complete removal of adaptor oligonucleotides. cDNA synthesis was monitored by incorporation of 50 µCi of [α-^{32}P]dCTP (Bresatec) at second strand synthesis, followed by electrophoresis of a sample on an alkaline agarose gel (McDonell *et al.*, 1977; Sambrook *et al.*, 1989) and subsequent autoradiography.

cDNA was ligated into *Eco*RI-digested and dephosphorylated λZAP®II vector (Stratagene) in accordance with the protocol supplied with the cDNA Synthesis Kit (Pharmacia). Phage DNA from a suitable ligation was packaged using Packagene® *in vitro* packaging extracts (Promega) and plated with *E. coli* XL1-Blue cells at a density of 1500 pfu per 150 mm plate. Prior to library screening, a random subset of recombinant plaques from the test plates were tested for cDNA inserts by phagemid rescue and *Eco*R1 restriction of the resulting plasmid. Phage were lifted in duplicate onto Hybond-N™+ membranes and differentially screened with ^{32}P-labelled single-stranded cDNA probes of total (Chapter 3) or poly(A)$^+$ (Chapter 4) RNAs from leaf and fibre.

Following secondary screening, recombinant pBluescript®SK(-) was rescued from λZAP®II using the ExAssist™/SOLR™ System, following manufacturer's instructions.

2.3.7 Rapid Amplification of cDNA Ends (5'-RACE)

Extension of the 5' ends of three fibre-specific cDNAs to obtain full-length cDNAs was achieved using the 5'-AmpliFINDER™ RACE kit (CLONTECH Laboratories) following manufacturer's instructions. Poly(A)$^+$ RNA (2 µg) isolated from 13 DPA fibres was converted into first-strand cDNA using a specific internal antisense primer designed from the cDNA sequence. After anchor ligation, PCR was carried out using another specific internal antisense primer and an anchor primer supplied by the manufacturer, together with Bresatec *Taq* DNA polymerase.

2.3.8 Phage Manipulation

2.3.8a Preparation of Plating Cells for Phage

Bacterial cells for plating of phage were prepared as per Stratagene instructions for the Predigested λZAP®II/*Eco*RI Cloning Kit and diluted to OD_{600} = 0.5 with 10 mM $MgSO_4$ before use.

2.3.8b Plating and Picking Phage

Phage and host cells were mixed and incubated at 37°C for 15-20 min prior to plating on the appropriate medium. Phage for the screening of libraries were plated on 150 mm plates using 600 µl of plating cells and 8 ml top agarose, melted to 48°C. Phage for secondary and tertiary screening, library titration or DNA preparations were plated on 85 mm plates using 200 µl of plating cells and 3 ml top agar or agarose, melted to 48°C. Phage from cDNA library test ligations were plated on 85 mm plates as above with the inclusion of 15 µl of 0.5M IPTG and 50 µl of 250 mg/ml X-gal to allow colour selection of recombinant phage. Plates were incubated at 37°C for between 8 and 16 hr.

Positive plaques which appeared on both the primary and duplicate lifts after autoradiography were aligned to the agar plates. Plaques were plugged into 0.5 ml SM buffer

and 20 µl chloroform using the large end of a sterile pasteur pipette and the phage allowed to elute at room temperature for 2 hr or at 4°C overnight. Phage were plated at low density on an 85 mm plate for secondary screening and the procedure repeated for a tertiary round of purification. Pure plaques were then plated to confluency on TB agarose plates for Lambda DNA preparation as described in Section 2.3.3e.

Genomic and cDNA libraries were titred using established protocols (Sambrook *et al.*, 1989) and an amplified stocks prepared according to Stratagene instructions for the Predigested λZAP®II/*Eco*RI Cloning Kit.

2.3.8c Mapping of Lambda Clones

Genomic clones were restriction mapped using the LambdaMap™ System (Promega), following manufacturer's instructions and utilising [γ-^{32}P]dATP (Bresatec) and T_4 polynucleotide kinase (Bresatec) to label the probes specific for each λ arm.

2.3.9 Construction of a Partial Genomic Library from Cotton

Cotton DNA (cv. Siokra 1-2) was partially digested with *Sau*3AI (Seed *et al.*, 1982) and size fractionated by sucrose gradient centrifugation as described by Ausubel *et al.* (1987) except that fractions were collected through a hole pierced at the base of the centrifuge tube in which the gradient had been established. Sucrose gradient fractions containing DNA fragments in the size range 15 kb to 23 kb (mean size 19 kb) were pooled and the DNA was ligated to *Bam*HI-digested, dephosphorylated λGEM-11™ arms (Promega) in pilot reactions (Sambrook *et al.*, 1989). The recombinant phage were packaged using Packagene® (Promega) and plated at high density with *E. coli* KW251 cells.

2.3.10 Transfer of DNA and RNA to Membranes

2.3.10a Southern Blotting

Following electrophoresis, DNA was transferred and alkali-fixed onto a Hybond™-N+

nylon membrane (Amersham, UK), following the instructions of the manufacturer.

2.3.10b Slotblotting of DNA for rRNA Gene Quantitation

Slotblotting of DNA samples onto pure nitrocellulose (Schleicher and Schuell) was achieved using a Minifold II Microsample Filtration Manifold (Schleicher and Schuell). DNA samples (1 µg) were made up to 40 µl with 1 x TE and denatured by the addition of an equal volume of 1M NaOH and incubation of the tubes at room temperature for 10 min. Samples were cooled on ice prior to the addition of 160 µl of ice-cold acid mix. Samples were loaded onto the slotblot under gentle vacuum and allowed to drain through, then washed with four applications of 50 µl of neutral mix. When the last wash had soaked through, the apparatus was disassembled, the membrane allowed to air-dry, and the DNA fixed onto the nitrocellulose by baking *in vacuo* at 80°C for 3 hr.

2.3.10c Slotblotting of RNA

Total RNA samples of approximately 5 µg were slotblotted onto a Hybond-N™+ membrane (Amersham) utilising the Microsample Filtration Manifold as above and following Amersham instructions for RNA dotblotting.

2.3.10d Plaque Lifting

Plaques were transferred onto Hybond-N™+ membrane (Amersham, UK) following the manufacturer's instructions. All lifts except secondary and tertiary screens of positive clones were done in duplicate. Briefly, the plates were chilled at 4°C for at least 2 hr. The filters were gently laid on the plaques and left at room temperature for 1 min and subsequent transfers of the same plate were left for an additional 30 sec. DNA was denatured, neutralised and alkali-fixed onto the filters as per Amersham instructions.

2.3.10e Colony Hybridisation

Colony hybridisation was achieved using the method described by Grunstein and

Hogness (1975). Recombinant bacterial colonies were patch plated onto an L-agar plate supplemented with 100 μg/ml ampicillin and replica plated onto a Hybond-N™+ membrane (Amersham) which overlayed another such plate. Following growth of bacterial colonies overnight at 37°C, the nylon filter was placed colony side up on a pad of three pieces of Whatman® 3MM paper soaked in lysing solution for 20 min, with blotting onto paper towel every 5 min. The filter was transferred to a fresh pad of 3MM paper soaked in denaturing solution plus 1% (w/v) SDS for a total of 20 min with blotting every 7 min. This procedure was repeated with 3MM soaked in neutralising solution before the filter was partially air-dried. Bacterial debris was removed from the filter using absorbent cotton wool and washing in 2 x SSC. The DNA was alkali-fixed onto the filters as per Amersham instructions.

2.3.10f Northern Blotting

Following electrophoresis of RNA, the gel was soaked in 0.05M NaOH, 1 x SSC for 20 min to partially hydrolyse the RNA, and the formaldehyde removed by subsequent soaking of the gel in 20 x SSC for 45 min. The RNA was then transferred onto Nytran®-plus (Schleicher and Schuell) polyamide membrane following the Amersham protocol for Hybond-N™+ and fixed onto the membrane by placing on Whatman® 3MM paper soaked in 0.05M NaOH for 5 min, followed by a brief rinse in 2 x SSC and storage at 4°C.

2.3.11 Radiolabelling of DNA probes

2.3.11a Oligolabelling Using Random Oligonucleotides

DNA fragments requiring radioactive labelling were purified by size fractionation on an agarose gel and freeze-squeezing as previously described. Labelling of 50-200 ηg of double-stranded DNA was performed by primer extension of random oligonucleotides (Feinberg and Vogelstein, 1983) using the GIGAprime DNA labelling kit (Bresatec) and [α-^{32}P]dATP (Bresatec or Amersham).

2.3.11b Labelling of Single-stranded cDNA Probes

To prepare single-stranded cDNA probes for differential screening, 2 µg to 4 µg of poly(A)$^+$ RNA in 3 µl of DEPC-treated H_2O was heated to 65°C for 5 min and transferred to ice for 5 min. After briefly microcentrifuging the sample, the following components were added in order: 1 µl DTT (10mM), 20 U RNasin® (Promega), 3.1 µl Tris-HCl pH 8.3 (400mM), 1.2 µl KCl (625mM), 0.5 µl $MgCl_2$ (500mM), 1.5 µl oligo(dT) (1.6 µg/µl, Amersham), 1.3 µl each of dATP, dTTP and dCTP (10mM), 120 µCi [α-^{32}P]dATP (Amersham) and 1 µl MLV reverse transcriptase (200 U/µl, Gibco BRL). The mix was incubated at 41°C for 2 hr.

2.3.11c End-labelling

Oligonucleotides for mapping of λGEM-11™ clones were end-labelled following the Promega protocol and utilising T_4 polynucleotide kinase (Boehringer Mannheim) and [γ-^{32}P]dATP (Amersham).

2.3.11d Labelling of "Cold" Clone cDNA Inserts

Inserts of "cold" cDNA clones were labelled by PCR, carried out on approximately 200 ηg of plasmid DNA with 3 µCi of [α-^{32}P]dATP (Bresatec), and using the reaction conditions outined in Section 2.3.15 for 45 cycles. The probe was denatured by incubation at 94°C for 2 min before cooling on ice and addition to a 0.5 ml hybridisation volume.

2.3.11e Removal of Unincorporated Nucleotides

Unincorporated nucleotides were removed from all labelling reactions except PCRs and end-labelled oligonucleotides by spun column chromatography through Bio-Gel P60 resin columns made from 100 µl G50-100 beads and 400 µl G100-200 beads (Bio-Rad). Columns were packed and equilibrated with 100 µl stop buffer by centrifugation at 170 x g for 2 min. Labelled DNA fragments were eluted by adding the labelling reaction plus stop buffer (to a

total volume of 100 µl) to the top of the column and recentrifugation at 170 x g for 2 min. Specific activity of the labelled DNA was determined by a scintillation count in ReadySolv™HP (Beckman).

2.3.12 Prehybridisation, Hybridisation and Washing

2.3.12a Prehybridisation Conditions

All filters were prehybridised, hybridised and washed under the same conditions except that plaque lifts were separated by nylon grids and prehybridised for at least 16 hr. Filters were prehybridised in a solution of 5 x SSPE, 5 x Denhardt's, 0.5% (w/v) SDS and 100 µg/ml denatured sonicated salmon sperm DNA at 65°C for a minimum of 2 hr. The volume of prehybridisation solution was approximately 0.1 ml/cm^2 of filter.

2.3.12b Probe Denaturation and Hybridisation

Following addition of 300 µg sonicated salmon sperm DNA, the radiolabelled probe was denatured by the addition of 34 µl 1M NaOH and incubation at room temperature for 10 min. The solution was neutralised by adding 34 µl 1M HCl and the denatured probe stored on ice. Filters were hybridised by labelled probe at a concentration of approximately 1 x 10^6 cpm per ml of hybridisation solution (to a maximum volume of 6 ml) for a minimum of 16 hr at 65°C. Oligo(dA) (Amersham) was added to hybridisation of cDNA library clones at a concentration of 10 µg per ml of hybridisation solution. Hybridisations were carried out at 65°C overnight.

2.3.12c Washing of Filters

Filters were washed in a solution of 2 x SSC, 0.1% SDS at 65°C for 30 min, 1 x SSC, 0.1% SDS at 65°C for 30 min and 0.5 x SSC, 0.1% SDS at 65°C for a further 30 min. Southern blots subject to a high stringency wash were incubated in 0.1 x SSC, 0.1% SDS at 65°C for 2 hr.

2.3.12d Autoradiography and Stripping of Filters

Filters were covered in plastic wrap to prevent drying out and exposed to X-ray film (Fuji RX) at -70°C for an appropriate period of time, backed by an intensifying screen (DuPont Hi-Plus). Slotblots used for rRNA gene quantitation were exposed at room temperature without an intensifying screen.

Southern blots and plaque lifts were stripped of probe by incubating for 10 min in 0.1 x SSC, 1% SDS at 90°C. Northern blots were stripped of probe by pouring boiling 0.1 x SSC onto them and allowing the solution to cool to room temperature, with mixing. All stripped filters were rinsed in 2 x SSC and stored in plastic wrap at 4°C for rehybridisation.

2.3.13 Design and Synthesis of DNA Oligonucleotides

All oligonucleotide primers were synthesised on 30ηM columns using an Oligo100 DNA Synthesiser (Beckman) and purified by the method of Sawadogo and Van Dyke (1991) using a DNA Ultrafast Cleavage and Deprotection kit (Beckman). Oligonucleotides for PCR were adjusted to a concentration of 10μM and primers for sequencing were diluted to 1μM. Annealing temperatures were calculated according to the formula $T_a = 2(A+T)+4(G+C)-5$ °C.

2.3.13a Primers for Sequencing

Oligonucleotide primers for sequencing were designed to unambiguous previously sequenced regions, priming towards a region of unknown sequence. They were generally synthesised as 20-mers at spacings of approximately 200 to 400 bp and contained a GC proportion of at least 50%.

2.3.13b Primers for PCR

Oligonucleotide primers for PCR were designed to contain the same proportion of purine and pyrimidine bases as that found in the target sequence. Efficient priming at high temperatures was maximised by the inclusion of two G or C residues at the 3' end of the

primers. In addition, primer pairs were checked at their 3' ends for the possibility of primer-dimer formation and intramolecular bonding using the Pcfold (Zuker, unpublished) and Molecule (Thompson, unpublished) computer programs. Primers used for 5'-RACE experiments were designed using the specifications of the manufacturer of the kit.

2.3.14 Sequencing of Double-stranded DNA Templates

2.3.14a Preparation of Template DNA

A 3 µg sample of the plasmid DNA to be sequenced was made up to 18 µl with dH_2O and 2 µl of 10 mg/ml RNaseA added before incubation at 37°C for 15 min. Following addition of 5 µl 1M NaOH, 1mM EDTA the solution was incubated for a further 15 min at 37°C. The denatured DNA was purified by centrifugation through a Sepharose CL-6B column (Murphy and Kavanagh, 1988), yielding 25 µl of single-stranded template DNA which was then stored on ice.

2.3.14b Sequencing Reactions

The dideoxy chain termination reaction (Sanger *et al.*, 1977) was performed using the Sequenase Version 2.0 kit (United States Biochemical Corp, Cleveland, Ohio) and [α^{32}P]dATP (Amersham, Bresatec), [α^{33}P]dATP (Amersham) or [α^{35}S]dATP (Amersham), as described by the manufacturer of the kit. Some sequencing was performed using the *fmol*™ DNA Sequencing System (Promega), following manufacturer's instructions. Where possible, sequence was confirmed by sequencing of both DNA strands. Compressed areas of GC-rich sequence were resolved using 7-deaza-guanosine in place of dGTP in the reaction mixes (Mizusawa *et al.*, 1986) and 1 µl each of 500 µM dATP and dTTP was added to sequencing reactions of fibre-specific cDNA clones.

2.3.14c Preparation of Sequencing Gels

A 4% (w/v) acrylamide, 8M urea, 1 x TBE gel solution was made up to 100 ml with

water and degassed for 10 min under vacuum. Following addition of 78 mg ammonium persulphate and 60 µl TEMED, the gel was mixed and poured between taped glass plates.

2.3.14d Denaturing Gel Electrophoresis

Sequencing reactions were run on a Model S2 sequencing gel electrophoresis system (BRL). Using 1 x TBE as the running buffer, the gel was pre-electrophoresed at 50 mA for 15 min. Following denaturation for 2 min at 90°C, 3 µl of each DNA sequencing reaction was loaded into the gel wells and electrophoresis continued until the tracking dye had reached the bottom of the gel plates. The gel was run between 30 mA and 50 mA but the voltage not allowed to exceed 1700 V. The loading process was repeated for medium- and short-length runs. Following electrophoresis, the sequencing gel was fixed in 20% (v/v) methanol and 10% (v/v) glacial acetic acid for 30 min to remove the urea. Finally, the gel was dried at 65°C for at least 3 hr and autoradiographed at room temperature.

2.3.14e Sequence Analysis Software

The compilation and comparison of nucleotide sequences from multiple cDNA clones and sequencing reactions was performed using various STADEN (Staden, 1980; Staden, 1984) programs and the GCG Sequence Analysis Software Package version 8 (Genetics Computer Group, Madison, WI, 1984; Devereux *et al.*, 1984). Multiple sequence alignments were done using the CLUSTALV program (Higgins and Sharp, 1988). Hydropathy profiles were obtained using the DNA Strider software package and BLASTN and BLASTX were used to search nucleotide and protein sequence databases, respectively (Altschul *et al.*, 1990; Gish and States, 1993).

2.3.15 Polymerase Chain Reaction (PCR)

Reaction mixtures (25 µl) for PCR contained 0.4 mM each of dATP, dCTP, dGTP and dTTP, 1 x PCR reaction buffer (Bresatec), 4 mM $MgCl_2$, 0.7-1.0 µM of each primer, 1.5 U of

Taq DNA polymerase (Bresatec) and template DNA. DNA was amplified in a programmable temperature controller (ARN Electronics) generally programmed as follows: initial template denaturation was at 94°C for 2 min followed by 25 PCR cycles of denaturation at 94°C for 1 min, primer annealing at 45°C for 1 min and extension at 72°C for 1 min. If the product was to be cloned, an additional cycle was carried out with an extension time of 7 min.

PCRs of plasmid and Lambda clones utilised approximately 200 ηg of DNA and amplification of cored "cold" cDNA clones utilised 12 µl of a 500 µl phage suspension. Annealing temperature varied depending on the primer pairs but was generally 45-50°C. The extension time also varied between different amplification reactions, according to the length of the fragment to be amplified (based on a synthesis rate of 1 kb per min).

2.3.16 Measurement of Cotton Fibre Lengths

Fibre lengths were measured using the method first described by Gipson and Ray (1969), beginning with 5 day old bolls and continuing at 5 day intervals until bolls were aged 50 DPA. Measurements were taken from 3-6 bolls per age and 6 seeds per boll, chosen at random except that clearly undersized or misshapen seeds were excluded from the analysis. After harvesting, the boll wall was carefully opened with a sharp knife and the intact locules floated in a beaker of dH_2O. For bolls 20 DPA and older, the locules were boiled for 2-5 min with 4 drops of HCl added per 100 ml of dH_2O to dissolve simple sugars and allow the seed with attached fibre to float free. After transfer to a beaker of cold water, six seeds were gently teased apart and each was placed on the convex side of a watchglass and the lint hairs made to stream out with a jet of water. All lint length measurements were made from the rounded side of the seed adjacent to the chalazal end, to the nearest 0.5 mm and using a centimetre rule.

2.3.17 Staining and Measurement of Cotton Fibre Nucleoli

Seeds with attached fibres were taken from midway along the carpel and fixed overnight in a solution of 3:1 methanol: glacial acetic acid before rinsing in 70% (v/v) ethanol and

storage under 70% (v/v) ethanol at 4°C until use. Fibres from the chalazal (rounded) end of each of four ovules were removed with forceps, blotted and spread onto glass slides. The fibres were stained with methyl green-pyroninY (Taft, 1951; Kurnick, 1952) for 10 min, a cover slip was placed onto the fibres and excess stain removed by even pressure. Nucleolar diameter was measured to the nearest 0.1 μm for 25 nucleoli per slide under using a light microscope set on 100 x objective and equipped with an OSM-D4 digital micrometer eyepiece (Olympus®).

2.3.18 Estimation of rRNA Gene Number

2.3.18a The Diphenylamine Reaction

The diphenylamine reaction for DNA quantitation was modified from the protocol of Burton (1956). For the DNA standard, calf thymus DNA (Sigma) was made to 100 μg/ml in 0.1 x SSC, filter-sterilised and frozen. The OD_{260} of the thawed solution was determined on the day of use and a dilution series constructed which ranged from 0 to 20 μg of DNA in 2.5 μg increments. DNA samples from each of four plants for each of six cotton varieties were adjusted to a concentration of 1 μg/μl (using UV spectrophotometry at a wavelength of 260 ηm) in order to keep all volumes to be pipetted in excess of 10 μl, thereby minimising errors.

DNA standards, together with 10 μg samples of the unknown DNAs, were made up to 100 μl with 0.1 x SSC. An equal volume of fresh 1M perchloric acid was added and the mixtures allowed to hydrolyse for 10 min at 70°C. Chilled, fresh acetaldehyde (1.6% (w/v) in dH_2O) and diphenylamine reagent were combined in a 1:20 ratio and 400 μl of the solution added to the DNA samples. Following thorough mixing and incubation overnight at 28°C, the colour changes which developed were read in triplicate by UV spectrophotometry at an absorption wavelength of 600 ηm. Average values for each of the standards were used to plot OD_{600} vs DNA amounts, and the standard curve was used to calculate the amount of DNA in the unknown samples by regression.

2.3.18b DNA Slotblotting for rDNA Quantitation

A known amount of standard rDNA (pGhR1) was used to construct two dilution series, containing 8, 10, 15, 20, 40 and 60 ηg of DNA. The dilution series, in duplicate, together with 1 µg samples of the unknown DNAs, as determined from the diphenylamine reaction and also in duplicate, were applied to a slotblot as outlined in Section 2.3.10b. Several measures were employed to reduce variation, including randomisation of samples to reduce position effects, loading of samples under slow vacuum to ensure even distribution of sample over slot and dilution of DNA samples to keep pipetted volumes greater than 20 µl.

2.3.18c Scanning Densitometry and Calculation of rDNA Copy Number

After autoradiography to obtain a preliminary determination of hybridisation, an appropriate exposure was quantified using an UltroScan XL Laser Densitometer (LKB, Sweden). A standard curve was calculated from the amounts of pGhR1 (ηg) vs the peak height measured (cm). Ribosomal gene copy numbers were calculated as shown in the following example except that numbers were not rounded off during calculation The peak height for Paymaster 909 for filter #1 and data set #1 was 9.07319 cm, which, from the regression of the standard curve for filter #1, data set #1 translates into 26.46 ηg of pGhR1 DNA. A sample of 1 ηg of pGhR1 DNA contains 0.76×10^8 copies of the 12 kb plasmid (each of which contains a single rDNA repeat unit). The 1 µg of Paymaster 909 DNA therefore contains 2.06×10^9 ($26.46 \times 0.76 \times 10^8$) copies of the rRNA gene repeat, which when divided by the number of genomes in 1 µg of DNA (2.63×10^5; Appendix 9), gives 7833 rDNA copies per Paymaster 909 tetraploid genome.

CHAPTER 3

Differential Screening of a Cotton Fibre cDNA Library Constructed from Total RNA

3.1 INTRODUCTION

The development of a cotton fibre is a complex process which requires the coordination of many macromolecular events. Chief among these is the biosynthesis of the fibre cell wall, the control of which is paramount in the determination of fibre characteristics. Although the major constituent of the cotton fibre is cellulose, other components may also be of importance. As previously discussed, there are many proteins, both structural and enzymatic, which could conceivably be up-regulated in a rapidly elongating cell such as the cotton fibre and their activity may influence the characteristics of the fibre. In addition, the fibre may contain unique forms of these proteins which relate to specific features of the cotton fibre cell wall.

Plant cell walls are specifically adapted to suit a particular cell type, and this is often reflected in the cell wall components and in cell wall architecture during expansion (Varner and Lin, 1989; McCann and Roberts, 1994). Genes involved in these adaptations must be regulated specifically. There are features of the cell wall of the developing cotton fibre which make this system unique. Mature cotton fibres lack lignin (Bacic *et al.*, 1988) but contain callose, a structural carbohydrate which is normally associated with stress response in higher plants, such as upon fungal infection or wounding (Maltby *et al.*, 1979). It is expected that cotton fibres contain further unique characters, controlled by genes which are expressed specifically in fibres but not in other cotton tissues. Such genes may have a substantial effect on the development of the fibre and on commercially-desirable characteristics.

The manipulation of cotton fibre quality requires an understanding of the genes and their protein products expressed in the differentiated cell. The promoters of fibre-specific genes clearly have potential in the genetic engineering of cotton fibres and will allow accurate tissue targeting of heterologous gene expression specifically to the fibres in transgenic plants. In addition, promoters which are active at different times in fibre growth can be used to drive expression of heterologous genes at a particular developmental stage.

Differential screening of a cotton fibre cDNA library was carried out in order to identify genes which are expressed exclusively or at elevated levels in the cotton fibres. Despite the prior isolation of several such genes (Section 1.6.2), the likely high complexity of the spectrum of proteins made in the cotton fibre indicate that there is scope for much overlapping work, especially on cotton varieties bred for Australian conditions.

3.2 Isolation and Preliminary Characterisation of Putative Fibre-specific cDNA Clones

Cotton tissues are not amenable to standard methods of RNA extraction, such as those of Chirgwin *et al.* (1979) and Cox (1968), due to the presence of polyphenolics and tannins in cotton cells (Katterman and Shattuck, 1983) which bind strongly to RNA and DNA upon cell lysis. Extraction of intact RNA from all cotton tissues therefore proved difficult and yields varied widely between different preparations. RNA extraction from fibres was especially difficult, most likely due to presence of the cell wall and in older fibres, presence of a secondary cell wall of cellulose.

RNA in relatively reliable yields was obtained from cotton fibres using a modification of the method of John (1992) and the integrity of the RNA was assessed using an *in vitro* coupled transcription-translation system, which produced a smear of radiolabelled protein products (not shown). Attempts to purify poly(A)$^+$ RNA from the total RNA pool using both messenger affinity paper and a streptavidin-based magnetic separation were unsuccessful, probably due to the low amount of starting material. Consequently, the cDNA library was constructed from total RNA, under the assumption that priming of cDNA synthesis with poly(T) would provide a reasonable selection for poly(A)$^+$ mRNAs. RNA was derived from fibres aged 10-24 DPA which covers a wide span of fibre development, including periods of both primary and secondary wall synthesis. Monitoring of cDNA synthesis by incorporation of [α–^{32}P]dCTP

yielded a discrete cDNA product, which was similar in size to that obtained with the use of control mRNA (Figure 3.1a). The band is formed as a by-product of the enzymology used to perform the Gubler and Hoffman (1983) method of cDNA synthesis, probably involving interactions between poly(A) fragments and oligo(dT) primers and its presence is indicative of a cDNA product.

Cloning of the cDNA product from cotton fibres aged 10-24 DPA resulted in a library consisting of 2.8 x 10^4 recombinant clones, which was differentially screened with labelled first-strand cDNA derived from 10-24 DPA fibre RNA and total leaf RNA. Leaf is a complex tissue which contains many different cell types. RNA samples from leaf should therefore contain transcripts which include the majority of so-called housekeeping genes expressed in a cotton plant. Use of the conventional technique of +/- hybridisation (Sargent, 1987) resulted in the isolation of four putative fibre-specific cDNA clones, together with six clones which appeared to hybridise more strongly to the fibre cDNA probe than to the leaf cDNA probe. In addition, many clones were identified which hybridised preferentially to leaf cDNA.

Excision of the inserts of the positive clones revealed the relatively short length of the inserted cDNAs, which ranged in size from approximately 0.1 kb to 1 kb (not shown). To investigate the expression patterns of the cDNA clones, RNA blot analysis was carried out using RNA from various cotton tissues. Hybridisation patterns were compared with those obtained using a control for RNA loading, the cotton rRNA gene cistron clone pGhR1 (Chapter 6; Figure 3.1c). The four putative fibre-specific cDNAs hybridised to transcripts present in leaf, flower and seedling RNA in addition to fibre RNA transcripts (Figure 3.1b). Erroneous identification of the cDNAs as fibre-specific probably arose as a result of a poor quality cDNA library which was plated at high density and required extended exposure after probing to identify positive plaques. In addition, the four putative fibre-specific cDNAs were located near the perimeter of the primary plaque lift and may have been omitted from the duplicate lift, which was then subject to probing with the leaf cDNA. Similarly, all six of the

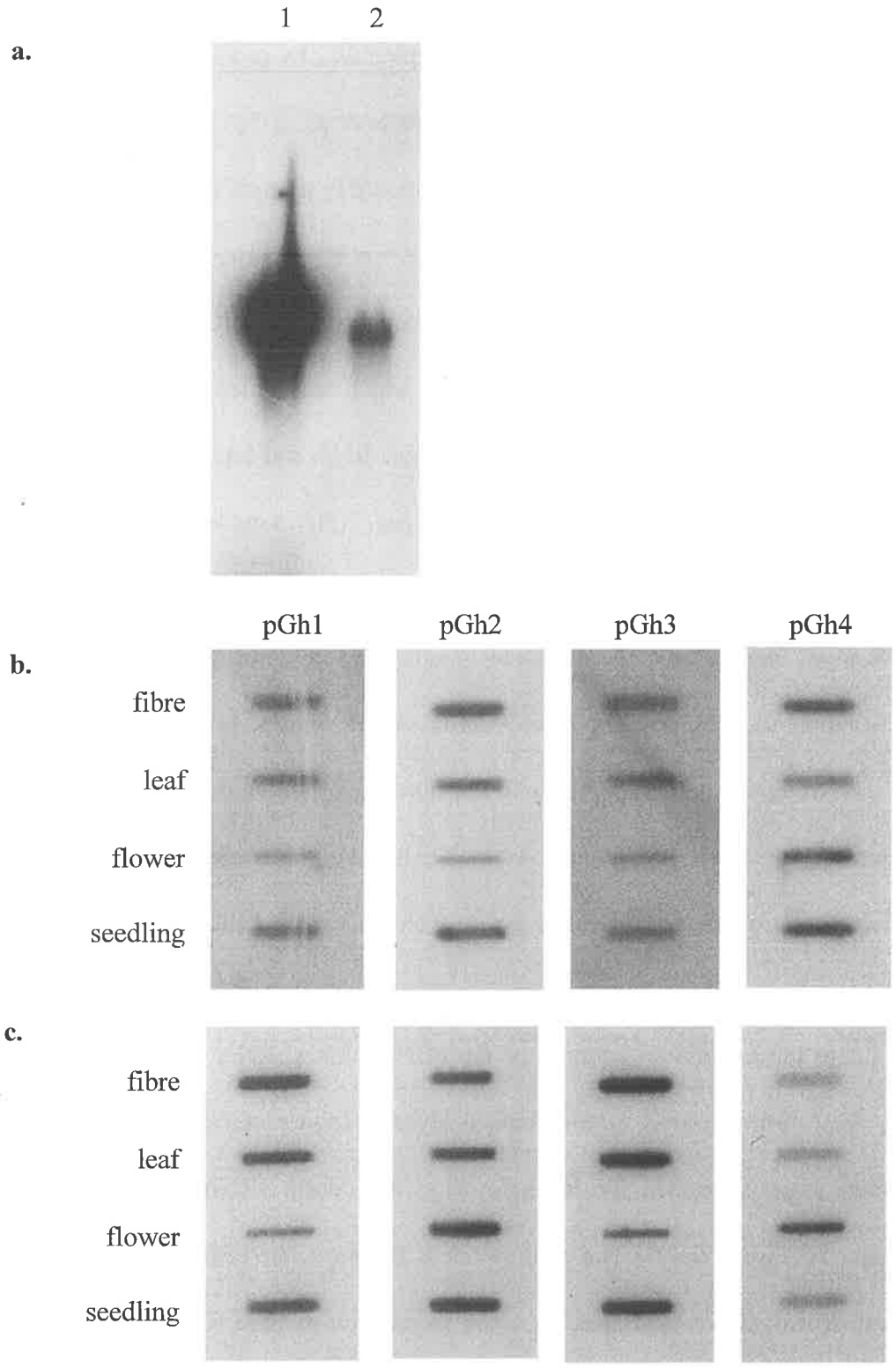


Figure 3.1: Analysis of putative fibre-specific cDNA clones

a. Alkaline gel electrophoresis of cotton fibre cDNA, synthesised from fibres aged 10-24 DPA (lane 1) together with a positive control in lane 2

b. RNA slotblot analysis of four putative fibre-specific cDNA clones. Each slot contains 5 μg of total RNA from fibre, leaf, flower or seedling tissue (indicated at left) and slotblots were hybridised with the specific probes as labelled.

c. RNA slotblots from part b., hybridised with the rDNA clone pGhR1.

clones which were identified on the basis of elevated hybridisation to fibre cDNA exhibited equal hybridisation to RNA from all cotton tissues (not shown). Hybridisation of an RNA slotblot with a probe derived from the amplification of pBluescript®SK(-) was negative after extended (3 days) exposure to X-ray film. However, one of the putative fibre-specific cDNAs, pGh4, appeared to exhibit elevated transcript levels in fibre RNA (Figure 3.1b) and was subject to sequence analysis.

Sequencing of the complete insert of clone pGh4 yielded 287 bp of sequence (Figure 3.2a) which showed significant similarity to the histone H4 gene from a variety of sources including tomato (Brandstädter *et al.*, 1994), sand urchin (Busslinger *et al.*, 1980), *Drosophila hydei* and yeast (Woudt *et al.*, 1983). The best match was 80.4% nucleotide identity to a 153 bp region of a histone H4 gene isolated from *Arabidopsis* (Chaboute *et al.*, 1987). A putative TAG termination codon was identified at nucleotide 152 of the pGh4 cDNA sequence (Figure 3.2a) and the subsequent sequence, representing the 3' untranslated region of the mRNA, exhibited substantially decreased similarity to the histone H4 homologues from other organisms. Accordingly, the 50 residue conceptual amino acid sequence of clone pGh4 showed significant similarity to the carboxy-terminal regions of the histone proteins mentioned. Histone proteins have a passive role in the packaging of DNA into nucleosome particles and may also play a crucial role in DNA replication and transcription (reviewed in Wolffe, 1991; Svaren and Hörz, 1993). High levels of histone H4 mRNA have been observed in proliferating tissues compared to mature tissues, and levels of H4 expression in particular cells are indicative of the distribution of mitotic activity (Brandstädter *et al.*, 1994). The synthesis of most histones is coupled to the cell cycle and occurs only during DNA replication. Since fibre cells are non-dividing, it is reasonable to expect that histone genes would be down-regulated in the fibres as compared to other cotton tissues. The reason for the observed pattern of expression is unknown but may be related to the dynamics of chromatin (Ausio, 1992) and the turnover of histone core proteins which occurs in all cells, including those not actively dividing.

a.

```
    GGAAACCCGTGGGGTTTTGAAGATCTTCTTGGAGAACGTGATTCGTGATGCAGTAACCTA
  1 ---------+---------+---------+---------+---------+---------+ 60
      E  T  R  G  V  L  K  I  F  L  E  N  V  I  R  D  A  V  T  Y

    CACCGAACATGCGAGGAGGAAGACTGTGACTGCTATGGATGTGGTTTATGCACTGAAGAG
 61 ---------+---------+---------+---------+---------+---------+ 120
      T  E  H  A  R  R  K  T  V  T  A  M  D  V  V  Y  A  L  K  R

    GCAGGGCAGGACATTGTATGGGTTTGGTGGTTAGATGATTTAGGGATTAGATCTTTCTAC
121 ---------+---------+---------+---------+---------+---------+ 180
      Q  G  R  T  L  Y  G  F  G  G  *

    ATGTTGGCGGATCTGGACTAGTTCTACATGATATCAAGAATGTTTAGGATTAGGAATAGG
181 ---------+---------+---------+---------+---------+---------+ 240

    GCTTGTAATTAAAGGGTGCTATTGNGATTAATTGGAATTTATAATTT
241 ---------+---------+---------+---------+------- 287
```

b.

```
    CTCACGGCCCCTCTTTTTGGACCCANNTCGGCCGGCCGACCGGGCGAAGACATTGTCAGG
  1 ---------+---------+---------+---------+---------+---------+ 60

    TGGGAAGTTTGGCTGGGGCGGCACATCTGTTAAAAGTACGCAGTGTCCTAAGATGAGCTC
 61 ---------+---------+---------+---------+---------+---------+ 120


    AACGAGAACAGAATCTCGTGTGGAACAAAAGGGTAAAAGCTCG
121 ---------+---------+---------+---------+--- 163
```

Figure 3.2: Sequence analysis of two putative fibre-specific cDNA clones

a. Nucleotide sequence of putative fibre-specific cDNA clone pGh4. The putative TAG termination codon at nucleotide 152 is indicated by an asterisk and the conceptual translation from nucleotide position 2 is shown in single letter code below the cDNA sequence.

b. Nucleotide sequence of putative fibre-specific cDNA clone pGh2.

Another putative fibre-specific cDNA clone, clone pGh2, was also sequenced. Sequencing of the entire insert of the cDNA clone yielded 163 bp (Figure 3.2b) which, when used to screen the nucleotide databases (GenBank and EMBL; January 1994), showed significant similarity to 26S rRNA genes from a variety of plants including lemon (Kolosha and Fodor, 1990), rice (Sugiura *et al.*, 1985), tomato (Kiss *et al.*, 1989) and rye (Appels *et al.*, 1986) with 89%, 87%, 86%, 85% nucleotide identity respectively. This explains the similar hybridisation patterns to cotton RNA obtained for clone pGh2 and pGhR1 (Figure 3.1b, c). The unexpected isolation of a ribosomal RNA sequence from the fibre cDNA library prompted the probing of a fraction of the library with the cotton rDNA clone pGhR1. The experiment revealed a high content of ribosomal sequences in the cDNA library, constituting approximately 80% of the hybridising plaques. The large proportion of ribosomal sequences can be attributed to non-specific priming of cDNA synthesis by poly(T), or the binding of poly(T) primer to A-rich regions of ribosomal RNAs. Approximately 80% of a total RNA sample is comprised of rRNA transcripts, with only 1-5% of the RNA attributable to mRNA species (Galau *et al.*, 1981).

3.3 Cold-plaque Screening and Sequence Analysis

The level of sensitivity of the differential screening method precludes the identification of clones derived from medium and low abundance class mRNAs. The lower limit of the differential screening technique is detection of abundant mRNAs which represent 0.1% of the total mRNA population (Sargent, 1987). A significant proportion (74%) of the recombinant cDNA clones in this differential screening experiment were "cold" plaques and failed to give a hybridisation signal with either the fibre or leaf cDNA probes even after extended exposure to X-ray film. As the library and fibre probe are derived from the same message population, the difference in hybridisation between recombinant plaques is likely to be due to variation in the corresponding transcript length or abundance. It is therefore assumed that many of these

cDNAs represent transcripts of low abundance in fibre cells. cDNAs derived from low abundance transcripts may be of interest since they might be expected to encode regulatory rather than structural proteins. Such a protein may have a range of regulatory functions related to development or differentiation of the fibre cell. In a screen for anther-specific clones in tobacco, Hodge *et al.* (1992) demonstrated that 44% of "cold" cDNAs were differentially expressed in the tobacco anther on diagnostic Northern blots but the same positive clones were undetected in a conventional differential screening experiment.

A total of 84 "cold" clones were selected at random from the 10-24 DPA fibre cDNA library and of these, 62 recombinant inserts were tested for hybridisation to the rDNA probe pGhR1 (Figure 3.3a, b). Again, PCR across the inserts revealed their generally small size, ranging from 100 bp to 1 kb (Figure 3.3a). A significant proportion (numbering 25) of the 62 recombinant clones hybridised to pGhR1 (Figure 3.3b) and were excluded from further analysis. The remaining 37 cDNA inserts were used as probes to RNA slotblots containing fibre, leaf, flower and seedling RNAs. No tissue-specific cDNAs were identified in the population of 37 cold clones screened but varying hybridisation patterns were observed compared with the positive control. A sample of these is shown in Figure 3.3c.

The clones which appeared to hybridise more strongly to fibre RNA than to RNA from other tissues, namely cold #6, cold #59 and cold #68, were chosen for sequence analysis. The inserts from two clones detected transcripts which were more abundant in leaf than in the other tissues and one of these, cold #27, was sequenced. A number of the clones with larger (300-500 bp) inserts were also sequenced and the results, which are summarised in Table 3.1, revealed that a high percentage of mitochondrial sequences were cloned. Sequences were derived from one DNA strand only and are largely incomplete except where indicated.

Table 3.1: Sequence analysis of cold cDNA clones

cDNA Clone	RNA Slotblot Analysis	Basepairs of Sequence	Sequence Similarity to GenBank and EMBL (January, 1994)
cold #3	even	343	mitochondrial 18S rRNA
cold #6	slight elevation in fibres	246	cytochrome b-5 (nuclear encoded)
cold #7	even	*354	no significant homology
cold #27	elevated in leaf	180	chloroplast 23S rRNA
cold #37	even	243	mitochondrial 26S rRNA
cold #38	even	129	mitochondrial 18S rRNA
cold #39	even	129	mitochondrial 18S rRNA
cold #42	even	*476	no significant homology
cold #54	even	314	mitochondrial gene *coxI* (cytochrome c oxidase subunit 1)
cold #56	?	*216	no significant homology
cold #59	slight elevation in fibres and flowers	*482	translational elongation factor 1α
cold #67	even	*345	no significant homology
cold #68	slight elevation in fibres	133	mitochondrial 26S rRNA
cold #71	even	*345	no significant homology

* Insert sequenced in full and in both directions

Of the 14 clones sequenced, five (cold #3, cold #37, cold #38, cold #39 and cold #68) contained inserts which showed significant sequence similarity to mitochondrial rRNA genes. The inserts from these clones hybridised equally to transcripts present in each of the four cotton tissues (Figure 3.3c). The clone which hybridised preferentially to RNA from leaf tissue, cold #27 (Figure 3.3c), appeared to encode a 23S rRNA from the chloroplast. While plastids are found in the cells of most tissues in a green plant, the major photosynthetic organ is the leaves and as such, leaf cells contain a large number of chloroplasts and therefore an abundance of plastid-derived rRNA transcripts. Fibre cells constitute a non-green tissue and have few, if any, chloroplasts (Ramsey and Berlin, 1976a). The cDNAs clones containing organellar

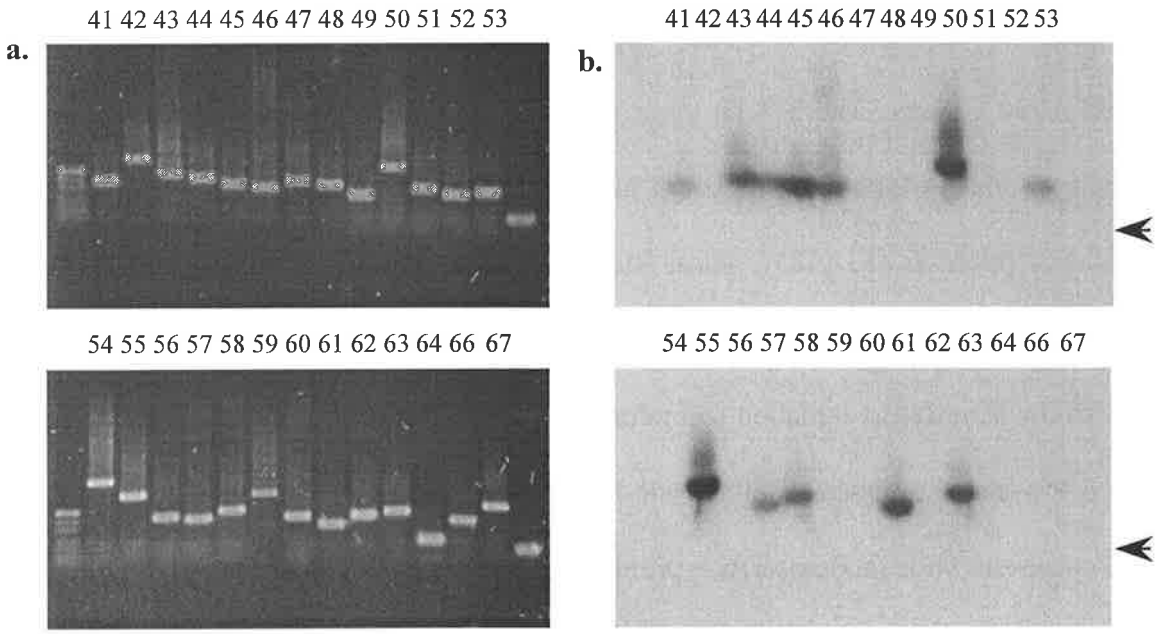
Figure 3.3: Analysis of "cold" cotton fibre cDNA clones

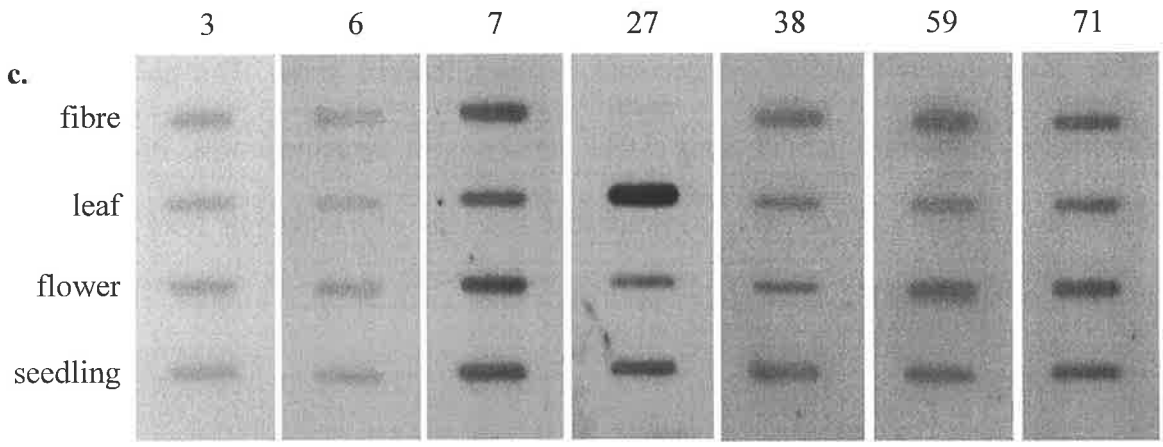
a. Agarose gel showing PCR products which resulted from amplification across the insert of a sample of 12 recombinant "cold" cDNA clones. Priming sites for T3 and T7, which flank the polylinker region of pBluescript®SK(-), were utilised in this experiment and each lane contains 10 µl of a 25 µl PCR reaction. Molecular weight markers (pUC19 DNA restricted with *Hpa*II) are shown at the left of the gel and the fragment sizes are 501 bp, 489 bp, 404 bp 331 bp, 242 bp, 190 bp, 147 bp and 111 bp. Numbers at the top of each lane represent the clone number, and the remaining lanes contain a 164 bp PCR product which results from use of pBluescript®SK(-) DNA as a template.

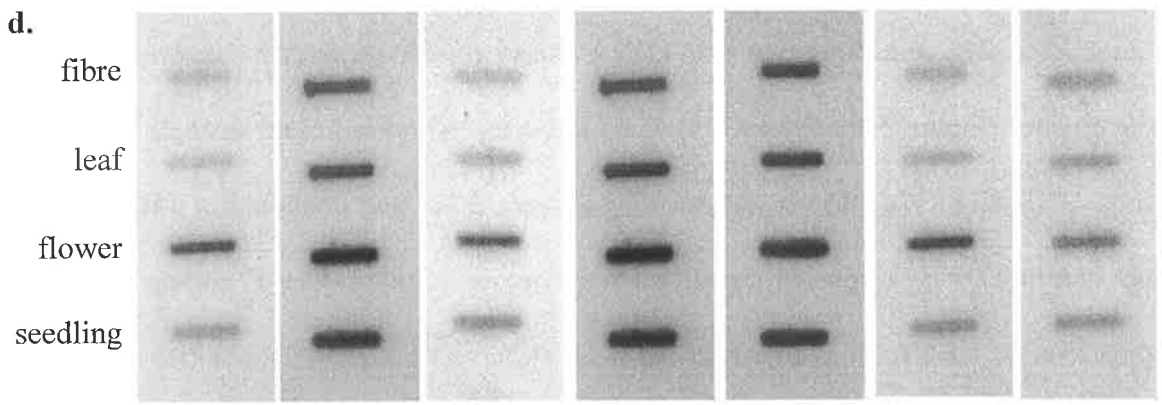
b. Hybridisation of a Southern blot of the above gel with the cloned cotton rRNA cistron pGhR1. Lanes are labelled as above and arrows indicate the position of the pBluescript®SK(-) control.

c. RNA slotblot analysis of a selection of "cold" fibre cDNA clones. Each slot contains 5 µg of total RNA from fibre, leaf, flower or seedling tissue (indicated at left) and slotblots were hybridised with the specific probes as labelled.

d. RNA slotblots from part b., hybridised with the rDNA clone pGhR1.


a.
41 42 43 44 45 46 47 48 49 50 51 52 53
54 55 56 57 58 59 60 61 62 63 64 66 67
b.
41 42 43 44 45 46 47 48 49 50 51 52 53
54 55 56 57 58 59 60 61 62 63 64 66 67


c.
3
6
7
27
38
59
71
fibre
leaf
flower
seedling


d.
fibre
leaf
flower
seedling

rRNA were not cross-hybridised by the nuclear rDNA clone pGhR1 (Figure 3.3b) and the sequence of the longest clone in each class of organellar transcript is presented in Figure 3.4a.

Another cold clone, #54, also appeared to encode a transcript derived from the mitochondrial genome. The 314 bp insert of cold #54 (Figure 3.4a) showed strong sequence similarity to the corresponding region of the cytochrome *c* oxidase subunit I gene (*coxI*) gene from maize (Isaac *et al.*, 1985), wheat (Bonen *et al.*, 1987), sugar beet (Senda *et al.*, 1991) and rice (Kadowaki *et al.*, 1989), with nucleotide identities 94%, 94%, 94% and 93% respectively. Translation initiation and termination codons were not identified in the sequence of cold #54 and the region of nucleotide homology occurred in the centre third of the *coxI* gene, consistent with random cDNA priming by poly(T). The conceptual translation of cold #54 (Figure 3.4a) produced an amino acid sequence which showed a high degree of similarity to coxI proteins. Cytochrome *c* oxidase is an enzyme complex of the inner mitochondrial membrane. Whilst some of the seven enzyme subunits are nuclear-encoded, in plants it seems that at least subunits I and II are encoded by mitochondrial genes. The predominance of mitochondrial transcripts in the fibre cDNA library is in agreement with observations by Ramsey and Berlin (1976a) who showed that the tips of elongating fibre cells contain numerous small vacuoles, lipid bodies and organelles such as dictyosomes, ribosomes and mitochondria.

Five of the cold clones that were sequenced contained inserts which did not show any sequence similarity to known sequences and encode as yet uncharacterised protein products (Figure 3.4b). The remaining two cold cDNA clones encoded transcripts derived from cotton nuclear genes (Figure 3.4c; Figure 3.5). Cold clone #6 hybridised more strongly to fibre RNA than to RNA from leaf, flower and seedling (Figure 3.3c) and contained a 246 bp insert, of which the first 79 bp displayed significant sequence similarity to the 3' coding region of the cytochrome b_5 mRNA from *Brassica oleracea* (Kearns *et al.*, 1992) and rice (Smith *et al.*, 1994), with nucleotide identities of 80% and 68% respectively. Although an open reading

frame was not identified in the sequence of cold #6, a putative TAG stop codon was identified at nucleotide position 77 and conceptual translation from nucleotide position 2 produced the 25 C-terminal residues of the cotton cytochrome b_5 protein (Figure 3.4c). Cytochrome b_5 is a well-conserved haem protein of around 130 amino acids which is associated with the endoplasmic reticulum of higher plants, animals and fungi (Ozols, 1989) and there are three published examples of plant cytochrome b_5 proteins, those from cauliflower (Kearns *et al.*, 1992), rice and tobacco (Smith *et al.*, 1994). Results from studies in tobacco suggest that there are a number of genes encoding plant cytochrome b_5 and that they may be differentially and/or developmentally regulated. The role of cytochrome b_5 in the microsomal membranes of plant cells is not well-understood, but in animals the protein functions as an electron donor in various oxidation/reduction reactions (Rich and Bendall, 1975). Why there seems to be a higher concentration of cytochrome b_5 transcripts in fibres than in other cotton tissues is uncertain.

The 482 bp insert of cold clone #59 was the longest cDNA clone sequenced (Figure 3.5) and was lacking in both a poly(A) tail and polyadenylation signal(s). No translation initiation codons were identified in the cold #59 insert sequence, but a putative TGA stop codon was identified at nucleotide position 292. Conceptual translation from nucleotide position 1 of the cDNA produced an amino acid sequence of 97 amino acids (Figure 3.5). A search of the databases (GenBank and EMBL; January, 1994) revealed that cold #59 showed significant sequence similarity to the eukaryotic elongation factor 1α (EF-1α). The cold #59 sequence, renamed pGhEF-1α, was similar to EF-1α genes from *Arabidopsis* (Curie *et al.*, 1992) and soybean (Aguilar *et al.*, 1991), and to EF-1α-related mRNAs isolated from tobacco (Zhu *et al.*, 1994), tomato (Pokalsky *et al.*, 1989) and maize (Berberich *et al.*, 1995), with 86%, 83%, 84%, 86% and 83% nucleotide identities respectively. Sequence similarity decreased markedly after the stop codon, to levels of 40% to 60%. At the amino acid level, the 97 residue conceptual translation of pGhEF-1α was strikingly similar to EF-1α proteins from *Arabidopsis*

Figure 3.4: Sequence analysis of "cold" cDNA clones

a. Sequences derived from organellar transcripts. Cold clone #3 is representative of the three clones, also including cold #38 and cold #39, which showed significant homology to the 18S rRNA from the mitochondria of various plants. Cold clone #37, together with cold #68, appeared to encode the 26S rRNA from the mitochondrion of fibre cells. Cold clone #27 encoded another rRNA transcript, the 23S rRNA derived from the chloroplast genome. The sequence of cold clone #54, encoding the mitochondrial coxI gene, includes the conceptual translation from nucleotide position 2 which is shown in single letter code below the cDNA sequence.

b. Sequences of the five "cold" clones which showed no significant similarity to known nucleotide sequences. Cold clones #67 and #71 were identical in both sequence and length and the sequence of one, cold #67, is shown here.

c. Nucleotide sequence of cold clone #6, which appears to encode cytochrome b_5. The putative TAG termination codon at nucleotide 77 is indicated by an asterisk and the conceptual translation from nucleotide position 2 is shown in single letter code below the cDNA sequence.

a.

cold #3: mitochondrial 18S rRNA

```
    CGGGTGCGTAACGCGTGGAATCTGCCGAACAGTTCGGGCCAAATCCTGAAGAAAGCTAAA
  1 ---------+---------+---------+---------+---------+---------+ 60

    AAGCGCTGTTTGATGAGCCTGCGTAGTATTAGGTAGTTGGTCAGGTAAAGGCTGACCAAG
 61 ---------+---------+---------+---------+---------+---------+ 120

    CCAATGATGCTTAGCTGGTCTTTTCGGATGATCAGCCACACTGGGACTGAGACACGGCCC
121 ---------+---------+---------+---------+---------+---------+ 180

    GGACTCCACGGGGGGCAGCAGTGGGGAATCTTGGACAATGGCGAAAGCCCGATCCAGCAA
181 ---------+---------+---------+---------+---------+---------+ 240

    TATCGCGTGAGTGAAGAAGGCAAGGCCCGCTCGTAAAGCTCTTTCGTCGAGTGCGCGATC
241 ---------+---------+---------+---------+---------+---------+ 300

    ATGACAGGACTCGAGGAAGAAGCCCCGGCTAACTCCGTGCCAG
301 ---------+---------+---------+---------+--- 343
```

cold #37: mitochondrial 26S rRNA

```
    GATCCGGGAGTCCCGTGTGGGGAAGGGCTCTCGCTCAACGGATCAAAGGTACGCCGGGGA
  1 ---------+---------+---------+---------+---------+---------+ 60

    TAACAGGCTGATGACTCCCAAGAGCTCTTATCGACGGAGTTTGGCACCTCGATGTCGACT
 61 ---------+---------+---------+---------+---------+---------+ 120

    CATCACATCCTGGGGTTGAAGAAGGTCCCAAGGGTTCGGTTNTNCNCCGATTCAAGTGGT
121 ---------+---------+---------+---------+---------+---------+ 180

    ACGTGAGTTGGGTTTAGAACNTCGTGAGCAAGTTCGGTTCCTAANCANCCGCAAATTCCT
181 ---------+---------+---------+---------+---------+---------+ 240

    GCA
241 --- 243
```

cold #27: chloroplast 23S rRNA

```
    CACCCCTACCTCCTTTATCACTGAGCGGTCATTTAGGGCCTTAGCTGGTGATCCGGGCTG
  1 ---------+---------+---------+---------+---------+---------+ 60

    TTTCCTCTCGACGATGAAGCTTATCCCCCATCGTCTCACTGGCCGACCTCGACCCCTGTT
 61 ---------+---------+---------+---------+---------+---------+ 120

    ATTTTGAGGTCATATCTAGTATTCAGAGTTTGCCTCGATTTGGTACCGCTCTACGCGCAT
121 ---------+---------+---------+---------+---------+---------+ 180
```

cold #54: mitochondrial *cox1* gene

```
    CGCGTCGCGCTGCCCGTGCCTACTTCACCGCAGCTACCATGATCATAGCTGTCCCACTGG
  1 ---------+---------+---------+---------+---------+---------+ 60
      A  S  R  C  P  C  L  L  H  R  S  Y  H  D  H  S  C  P  T  G

    AATCAAAATCTTTAGTTGGATCGCTACCATGTGGGGAGGTTCGATACAATACAAAACACC
 61 ---------+---------+---------+---------+---------+---------+ 120
      I  K  I  F  S  W  I  A  T  M  W  G  G  S  I  Q  Y  K  T  P

    CATGTTATTTGCTGTAGGGTTCATCTTTTTNTTCACCATAGGGGGACTCACTGGAATAGT
121 ---------+---------+---------+---------+---------+---------+ 180
      M  L  F  A  V  G  F  I  F  ?  F  T  I  G  G  L  T  G  I  V

    CCTGGCAAATTCTGGGCTAGACATTGCTCTACATGATACTTATTATGTGGTGCACATTTC
181 ---------+---------+---------+---------+---------+---------+ 240
      L  A  N  S  G  L  D  I  A  L  H  D  T  Y  Y  V  V  H  I  S

    CATTATGTACTTTCTATGGGAGCCGTTTTTAGCTTTATTTGCAGGATTTTACTATTGGGT
241 ---------+---------+---------+---------+---------+---------+ 300
      I  M  Y  F  L  W  E  P  F  L  A  L  F  A  G  F  Y  Y  W  V

    GGGTAAAATCTTTA
301 ---------+---- 314
      G  K  I  F
```

b.
cold #7

```
    TCAGAGTATGCAGACGCCACCGGGCAGAATGAAGAGGGTCTTGACGAGATTGTGGACGCA
  1 ---------+---------+---------+---------+---------+---------+ 60

    ATGTCACCGGAAATGGCACCAGCAAGTTCACCCACAATGGCATTTGAAGGGTCTTTTGGA
 61 ---------+---------+---------+---------+---------+---------+ 120

    CCCGGACCAGCTATGACCATTGAAGACGGTGCGGCGGAGGGAGAGGGAGAGGCAGAGGCA
121 ---------+---------+---------+---------+---------+---------+ 180

    GAGGGAGCAACAGAAAGTTAGATACGCAGTACCAGTAAATTTAAATAAGGATCTCGATCC
181 ---------+---------+---------+---------+---------+---------+ 240

    ATATGTTCGTTATCTTACTATTTTGGTATTTCATCCCTCATAAAAGAAAAGTGAACCATA
241 ---------+---------+---------+---------+---------+---------+ 300

    ACTGTAAAGTTCACGACCATAAGTAATATATTTATTTATTTACCTGCCCACCCA
301 ---------+---------+---------+---------+---------+---- 354
```

cold #42

```
    GAATCTTACAAGGCATGGGCAGCTTTGGAGCTAGAGGAAGACAAACAAGTGAAAGGAGCT
  1 ---------+---------+---------+---------+---------+---------+ 60

    GAAATAACCATGCAACAAGCAGAGGATTATTTTGACTCGGTCATGGAAACCGCCGTGGAT
 61 ---------+---------+---------+---------+---------+---------+ 120
```

```
    GAGTTCAGGCGATTCGAGGAAGAAATGGAACGCGAGTCAAAAGCTGAACTCAGTGGTGTT
121 ---------+---------+---------+---------+---------+---------+ 180

    GATGATACTGCTGAGACAGTTAAAAAAATGGGGGATTTGATGGAAAAAGGTGCAAACATT
181 ---------+---------+---------+---------+---------+---------+ 240

    GCGTCTAAATTGTATGTTGAAGCTGCTATGAAATCTGCAGGTTTTAATGGACTTTCACCT
241 ---------+---------+---------+---------+---------+---------+ 300

    AACAAGGTTCATCCTTCATGAAAAAAGTTATATATATATTATATTTATTTAATTTCATTT
301 ---------+---------+---------+---------+---------+---------+ 360

    TTTTGTAATTTATATGTATGATTGATAAGGAACCTTGATGGGGGTTTCTTTGATGTTACT
361 ---------+---------+---------+---------+---------+---------+ 420

    ATTTTTTTTAATGGATTATTTGTTTTTAATTAAAATCCATGAACTGGGTTTCAAGA
421 ---------+---------+---------+---------+---------+------ 476
```

cold #56

```
    CTCACATATATTATTATTAACATCAAGCAGCGCGCAGAAGCAAAATCAACAAAAAAATTC
  1 ---------+---------+---------+---------+---------+---------+ 60

    TGGAAATACAAATAATGTGGATAACAAGAATATGTTCAACGACCAATCAGACCCTTATCG
 61 ---------+---------+---------+---------+---------+---------+ 120

    ATCCATTTGAAAAACGAAAATTTATAGTAAAAAATGCAGACAAACTTCATATATATCCGC
121 ---------+---------+---------+---------+---------+---------+ 180

    TCATGAGAGAAATCCAATGCAAGCCAATGCAAGCCA
181 ---------+---------+---------+------ 216
```

cold #67

```
    CGGGAAATGGTCCCGTGTTCAGGCTGAAGGAGACTATAAAGCAAAACTCGGAATCACAAT
  1 ---------+---------+---------+---------+---------+---------+ 60

    TCTTGGAGAAGCAGTGACTGTTGAAGATGAAGTCGTGGTCACCAATAGCATTGTTCTTCC
 61 ---------+---------+---------+---------+---------+---------+ 120

    GCACAAGACGTTAAACGTCAGTGTTCAAGACGAAATACTATTATGATATTTGTATCCGCT
121 ---------+---------+---------+---------+---------+---------+ 180

    CCCGTTTGCCGCCTCTAGGCTCCCTCCTATTTCGCAGCCCCTTTATCTACCCTAGTTGTG
181 ---------+---------+---------+---------+---------+---------+ 240

    ATGTAATGGTTACTGTGCATTTTGTGAACTAGGGAAATCAGTAGTATTTCTTAATTTTTA
241 ---------+---------+---------+---------+---------+---------+ 300

    CTCGTATGTATATTGAGATGCTCAAATTAGAATAAAATTTCAACA
301 ---------+---------+---------+---------+----- 345
```

c.

cold #6: nuclear-encoded cytochrome b_5 gene

```
    CAAGATCCTACAATTTCTTGTTCCCCTCTTGATCCTGGGCTTGGCTATTGCGGTACGTCA
  1 ---------+---------+---------+---------+---------+---------+ 60
      K  I  L  Q  F  L  V  P  L  L  I  L  G  L  A  I  A  V  R  H

    CTATACCAAGAAAGAGTAGCTTATTACTATACAGGACTTTACCTGTTACGAGGGAAGAAT
 61 ---------+---------+---------+---------+---------+---------+ 120
      Y  T  K  K  E  *

    CAGTGTGTGTCTCATCTGTTAGTTGAAATAGCACAGTCATATCTTTTGTTTAATTTCCTT
121 ---------+---------+---------+---------+---------+---------+ 180

    TTAGCATTTGTGTGTCGAACTAAATACCTGTTGATGATAGCTCCATTGTCAAAAGACAAT
181 ---------+---------+---------+---------+---------+---------+ 240

    ATATTT
241 ------ 246
```

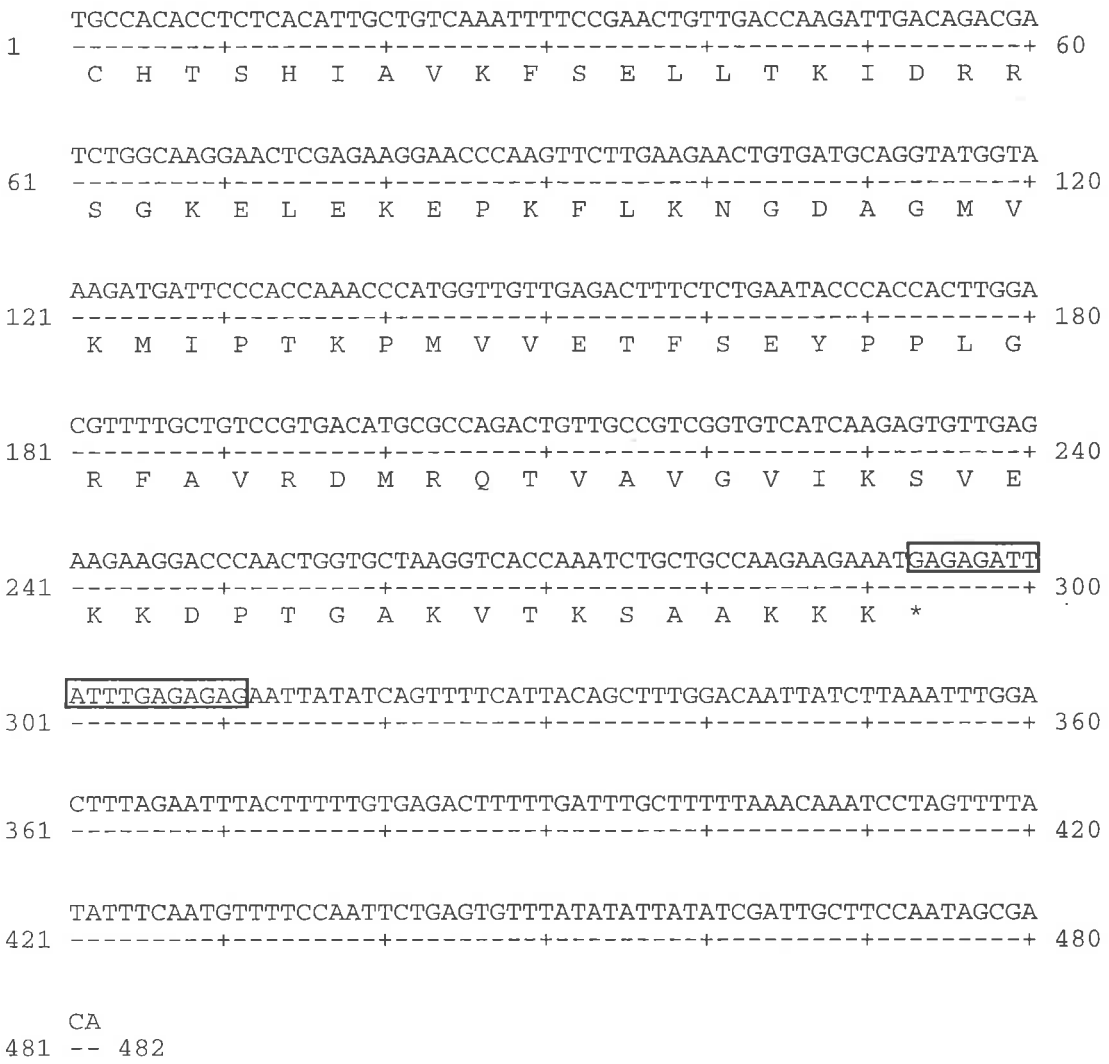


Figure 3.5: Sequence of the cotton EF-1α cDNA clone pGhEF-1α

Nucleotide sequence of the cotton EF-1α cDNA, pGhEF-1α (cold #59). The deduced amino acid sequence, translated from nucleotide position 1, is shown in single letter code below the DNA sequence and the putative stop codon at nucleotide position 292 is indicated by an asterisk. The region which corresponds to the oligonucleotide primer EFP1 is boxed.

(92% identity; Axelos *et al.*, 1989), tobacco (91% identity; Zhu *et al.*, 1994), soybean (92% identity; Aguilar *et al.*, 1991) and tomato (90% identity; Pokalsky *et al.*, 1989) together with EF-1α proteins isolated from monocots such as wheat (88% identity; Metz *et al.*, 1992) and from divergent species such as *Xenopus* (74% identity; Djé *et al.*, 1990), human (72% identity; Brands *et al.*, 1986), yeast (73% identity; Nagata *et al.*, 1984) and *Drosophila* (66% identity; Hovemann *et al.*, 1988) (Figure 3.6).

Notably, none of the cDNA clones sequenced from the 10-24 DPA fibre cDNA library contained poly(A) tails and this contrasts with the finding that, in plants, all mRNAs that have been examined are polyadenylated, including histone mRNAs. The only known poly(A)$^-$ mRNAs present in plant cells are viral in origin (Gallie and Kobayashi, 1994). One explanation is that the truncated cDNAs result from inefficient synthesis of the second cDNA strand such that the product terminates prior to the poly(A) tail. A less likely explanation is that the cDNA products represent mRNA degradation products, since poly(A) tail removal constitutes one of the earliest events in mRNA degradation (Thompson *et al.*, 1992).

3.4 The Eukaryotic Elongation Factor 1α

3.4.1 Role of EF-1α in Translation

The process of translation of genetic information from mRNA to protein follows a distinct pathway. One step, the elongation of the amino acid chain, involves a series of protein components that have been classified according to their function as elongation factor 1 (EF-1) and elongation factor 2 (EF-2). In eukaryotes, the cytoplasmic EF-1 complex is thought to consist of four different subunits, EF-1α, EF-1β, EF-1γ and EF-1δ. EF-1α is functionally homologous to the bacterial factor EF-Tu which catalyses the GTP-dependent binding of the aminoacyl tRNA to acceptor site A on the ribosome (Riis *et al.*, 1990). This process involves the concomitant binding and hydrolysis of GTP, the binding of aminoacyl tRNA and the

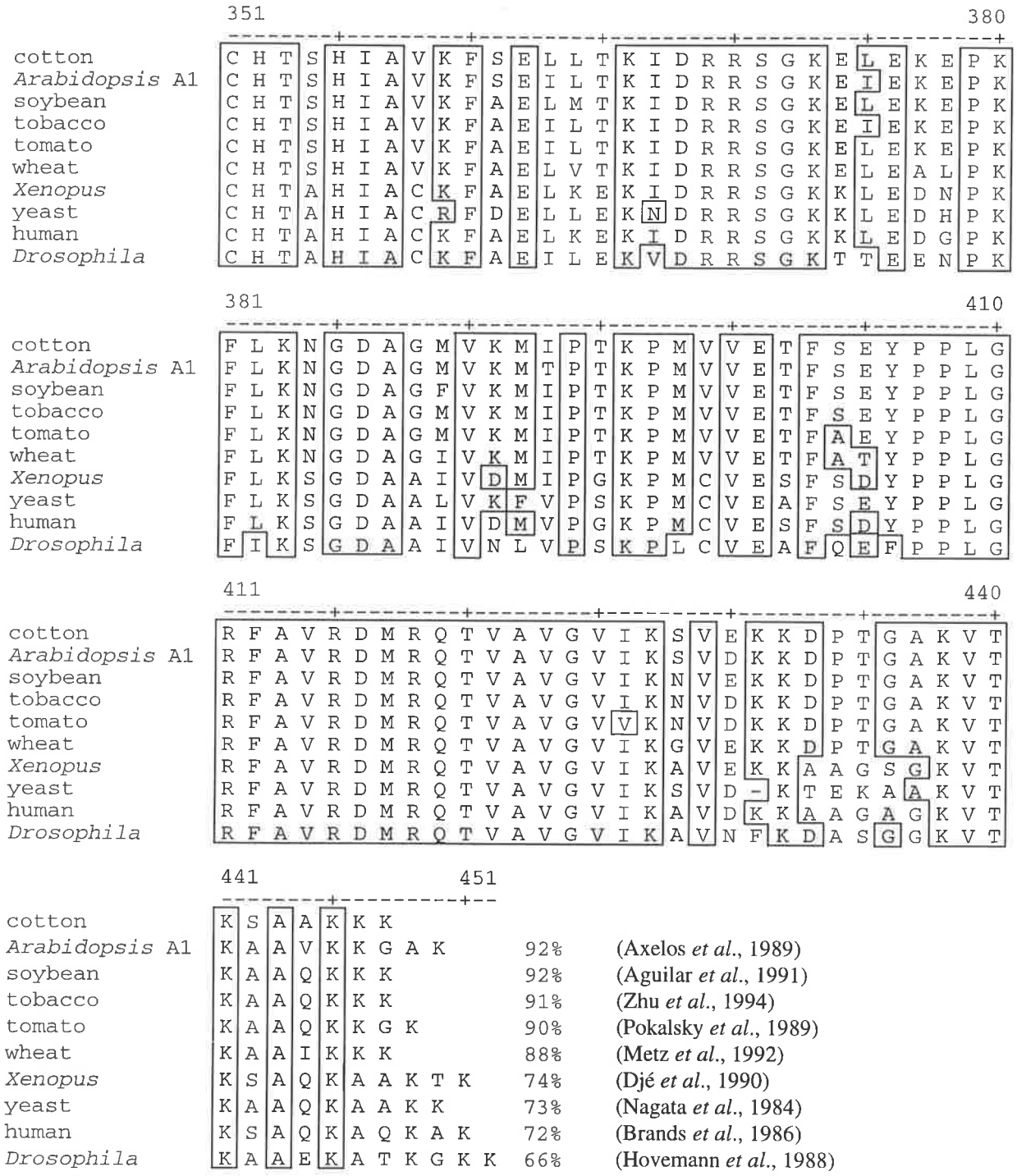


Figure 3.6: Alignment of elongation factor 1α sequences

Sequence comparison between the conceptual translation of pGhEF-1α and the C-termini of nine EF-1α proteins from a selection of other plants and eukaryotes. Boxes indicate homologous amino acids in at least seven of the sequences, including that from cotton, and homology with the cotton fibre EF-1α is indicated at the end of each sequence. Numbers refer to residues in the five plant EF-1α proteins.

recognition and interaction with the ribosome. The multifunctional demand on EF-1α clearly leaves little room for evolution of divergent protein structures and the striking degree of homology between EF-1α proteins (Figure 3.6) is interpreted as a conservation of active sites. Some of the most highly conserved regions have been correlated with functional domains. The most stringently conserved parts of EF-1α occur at the N-terminus and have been assigned to GTP binding activity, while other regions of the protein show homology to several classes of nucleotide-binding proteins and the prokaryotic elongation factors.

Several studies have correlated the abundance of translation factors, particularly EF-1α, with the efficiency of translation. A close relationship between the level of mRNA for EF-1α and the rate of protein synthesis has also been reported (Pokalsky *et al.*, 1989; Ursin *et al.*, 1991; Kawahara *et al.*, 1992; Curie *et al.*, 1993; Morelli *et al.*, 1994) and expression of EF-1α is the major limiting step in polypeptide elongation (Riis *et al.*, 1990). Increased amounts of EF-1α activity are generally found in developing tissues compared to mature tissues and the expression of EF-1α may serve as an indicator of translational stimulation. Further, EF-1α transcription is induced by hormone application (Ursin *et al.*, 1991), wounding (Morelli *et al.*, 1994), cold stress (Dunn *et al.*, 1993; Berberich *et al.*, 1995), hypoxic stress (Vayda *et al.*, 1995), exposure to light (Aguilar *et al.*, 1991) and other treatments that increase protein synthesis. Conversely, the large decreases in translational activity that occur upon starvation, fungal infection or senescence have been attributed to depletion of elongation factors or correlated with decreased expression of EF-1α (Webster, 1985; Mahe *et al.*, 1992).

3.4.2 Alternative Roles for EF-1α in Plant Cells

Besides its role in polypeptide elongation, recent evidence suggests that EF-1α is involved in a variety of biological processes and may be regarded as a multifunctional protein (Durso and Cyr, 1994b). For example, a vitronectin-like adhesion protein isolated from tobacco (Zhu *et al.*, 1994) has been implicated in cellular processes such as plasma membrane-

cell wall adhesion, pollen tube extension, and bacterial infection and is localised in the cell wall but is structurally and immunologically related to EF-1α. Numerous studies have demonstrated the binding of EF-1α to actin filaments and microtubules *in vivo* and *in vitro* (Condeelis, 1995). Such EF-1α/cytoskeletal interactions have prompted the hypothesis that EF-1α is involved in feedback regulation of cytoskeletal protein synthesis (Durso and Cyr, 1994a). Furthermore, the emerging role of EF-1α in intracellular signalling (reviewed in Riis *et al.*, 1990) has wide-ranging implications for the coordination of various cell functions, cell division, differentiation, aging and transformation.

3.4.3 Expression of EF-1α Genes

In eukaryotes, EF-1α is the second most abundant protein after actin, constituting 1-2% of the total protein in normal growing cells (Slobin, 1980). The regulation of EF-1α is of interest because it is an example of a protein which is directly required for the synthesis of all cellular proteins and appears to be essential for cell viability (Cottrelle *et al.*, 1985). In addition, information on the mechanism of regulation of housekeeping genes in higher plants is scarce. It is not clear to what extent EF-1α activity is controlled at the level of mRNA or protein as examples of transcriptional, translational and post-translational controls have been seen for EF-1α in a variety of organisms.

The promoters of EF-1α genes are extremely strong and may therefore have applications in transgenics and genetic engineering. Strong promoters have been demonstrated for EF-1α genes from human (Kim *et al.*, 1990) and fungal systems (Steiner and Philippsen, 1994). The A1 gene for EF-1α in *Arabidopsis thaliana* has a promoter which is extremely strong and which mediates a transient expression about two-fold higher than that obtained using the CaMV 35S promoter (Axelos *et al.*, 1989). The latter is one of the strongest plant promoters known, in both transient expression systems and stably transformed plant cells (Odell *et al.*, 1988). Dissection of the A1 promoter has identified several *cis*-acting regulatory elements,

which are located both upstream and downstream of the site of transcription initiation (Curie *et al.*, 1991; Curie *et al.*, 1992; Curie *et al.*, 1993), and two *trans*-acting protein factors are required for maximal expression of the gene in transfected *Arabidopsis* cells. The high GUS activity mediated by the *Arabidopsis* A1 promoter was found to be conserved among several angiosperms, suggestive of a universal control mechanism for plant EF-1α genes. A domain termed the TEF1 box is shared by the promoter of an EF-1α gene from tomato and the four EF-1α genes isolated from *Arabidopsis* (Curie *et al.*, 1991; Curie *et al.*, 1992).

Although EF-1α is involved in so-called housekeeping functions, transcription of some EF-1α genes is differentially and developmentally regulated, such as during somatic embryogenesis of carrot cells (Kawahara *et al.*, 1994). Members of the EF-1α gene family in *Xenopus* and *Drosophila* are also regulated temporally and spatially during development (Hovemann *et al.*, 1988; Djé *et al.*, 1990), which supports the notion of additional roles for EF-1α.

3.4.4 The Cotton Fibre EF-1α

The cotton fibre EF-1α is of particular interest because of its role in the control of translation. For genetic engineering which involves manipulation of gene expression to be successful, the level of control of expression for the particular gene must be understood. For example, onset of secondary wall synthesis in the cotton fibre may trigger the expression of a particular set of genes at the translational level (Alexander *et al.*, 1983). Proteins such as EF-1α may be pivotal to such a developmental switch and in turn, the expression of these genes may be under transcriptional control and therefore more amenable to control by genetic engineering. EF-1α may be up-regulated in fibre cells and flower tissue (Figure 3.3c) but the result is unclear and insufficient RNA material prevented Northern analysis and elucidation of the temporal pattern of EF-1α gene expression in developing cotton fibres. The high numbers of ribosomes and enlarged nucleoli observed in elongating fibre cells compared with adjacent

epidermal cells (Ramsey and Berlin, 1976a; Ramsey and Berlin, 1976b) suggest an increased capacity for protein synthesis, which may be concurrent with a general increase in translation machinery components. A tissue such as cotton fibres which are particularly active in protein production might therefore be expected to have increased levels of elongation factor proteins.

3.5 Southern Analysis of EF-1α in the Cotton Genome

The copy number of EF-1α-related sequences in Siokra 1-2 genomic DNA was estimated by a Southern hybridisation experiment. With each restriction enzyme used, at least two major fragments and a variable number (2-10) of fainter bands hybridised to the pGhEF-1α cDNA insert (Figure 3.7). The latter represent more divergent members of the gene family. A large number of EF-1α pseudogenes have been reported in human, murine, porcine and avian species (Madsen *et al.*, 1990). In particular, the *Eco*RI restriction contained many hybridising bands (Figure 3.7, lane 1). While it is unknown whether the full-length EF-1α cDNA from cotton contains an *Eco*RI site, the EF-1α genes from other plants which have been characterised to date lack *Eco*RI sites, both in their coding and non-coding sequences.

Hybridisation to several DNA fragments in each lane suggests that the isolated EF-1α cDNA clone belongs to a family of related sequences in the cotton genome, which is in accord with findings in other plant species. EF-1α is encoded by at least four related genes in *Arabidopsis* (Axelos *et al.*, 1989) and carrot (Kawahara *et al.*, 1992), four to eight genes in tomato (Pokalsky *et al.*, 1989; Shewmaker *et al.*, 1990) and at least six family members in maize (Berberich *et al.*, 1995).

3.6 Isolation of Chromosomal Genes for EF-1α from Siokra 1-2

Verification of the above result came from screening of a genomic library for EF-1α clones. Hybridisation of a *G. hirsutum* cv. Siokra 1-2 partial genomic library consisting of

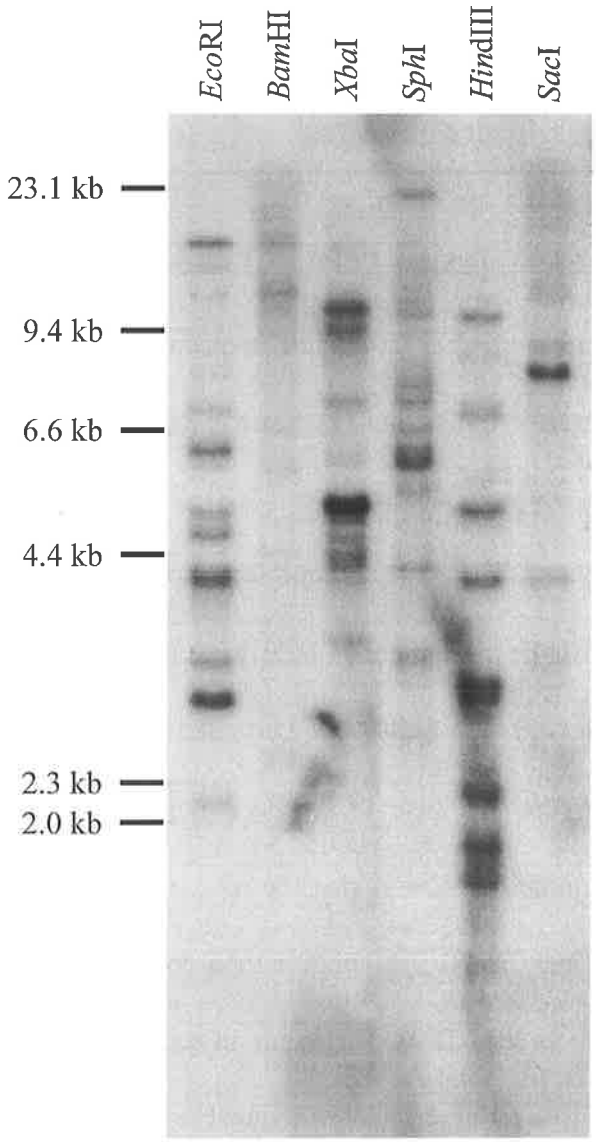


Figure 3.7: Southern analysis of cDNA clone pGhEF-1α

Genomic Southern blot probed with the pGhEF-1α insert. Each lane was loaded with 5 μg of total genomic DNA cut with enzymes which recognise 6 bp target sites. Numbers on the left represent the molecular weights of wild type Lambda DNA restricted with *Hin*dIII (kb) and lanes are *G. hirsutum* L., cv. Siokra 1-2 DNA restricted with the endonucleases indicated.

1.5 x 10^5 recombinants with pGhEF-1α identified 61 positive clones (not shown). The clones could be classified into three groups according to the intensity of hybridisation signal with EF-1α cDNA. Six distinct clones, two of which gave a signal of low intensity and four of which gave a signal of high intensity were purified and subject to further analysis. Restriction of recombinant phage DNA with *Xho*I and *Eco*RI, combined with Southern blot hybridisation using the cDNA insert of pGhEF-1α as a probe, revealed that different members of the EF-1α gene family had been isolated (Figure 3.8a, b). The restriction enzyme *Xho*I excised the clone inserts which ranged in size from 15 kb (λL4) to 20 kb (λL3). Use of pGhEF-1α as a probe to the genomic clones produced one hybridising DNA fragment per clone (Figure 3.8b) except for λS5, in which two *Xho*I fragments hybridised weakly to the cDNA probe. As expected, hybridisation was strongest to fragments in clones λL2, λL3 and λL4, and weaker in clones λS3 and λS5. Relatively weak hybridisation to λL5 is probably a reflection of DNA loadings on the agarose gel (Figure 3.8a). The six distinct clones represent four of the *Eco*RI bands observed on the genomic Southern. Genomic DNA fragments of approximately 6 kb, 4 kb, 11.5 kb and 3 kb (Figure 3.7, lane 1) correspond to the hybridising *Eco*RI fragments of genomic clones λL3, λL4, λS3 and λS5 respectively.

To ascertain which of the genomic clones, if any, contains the gene which encodes pGhEF-1α, the genomic clone Southern was probed with the 3' untranslated region of the transcript, since it is in this region that EF-1α gene family members vary (Axelos *et al.*, 1989). In a previous experiment, a probe specific for the 3' end of a tomato cDNA differentiated between transcripts derived from different members of the EF-1α gene family (Shewmaker *et al.*, 1990). Lack of appropriate restriction sites necessitated the design of a PCR primer, EFP1, located 171 bp from the 3' terminus of the pGhEF-1α cDNA (Figure 3.5; Table 2.3b). Use of EFP1 in combination with pBluescript®SK(-) primer T7 amplified a 276 bp fragment of pGhEF-1α (not shown). Restriction with *Eco*RI yielded a 188 bp fragment which

Figure 3.8: Southern analysis of EF-1α genomic clones

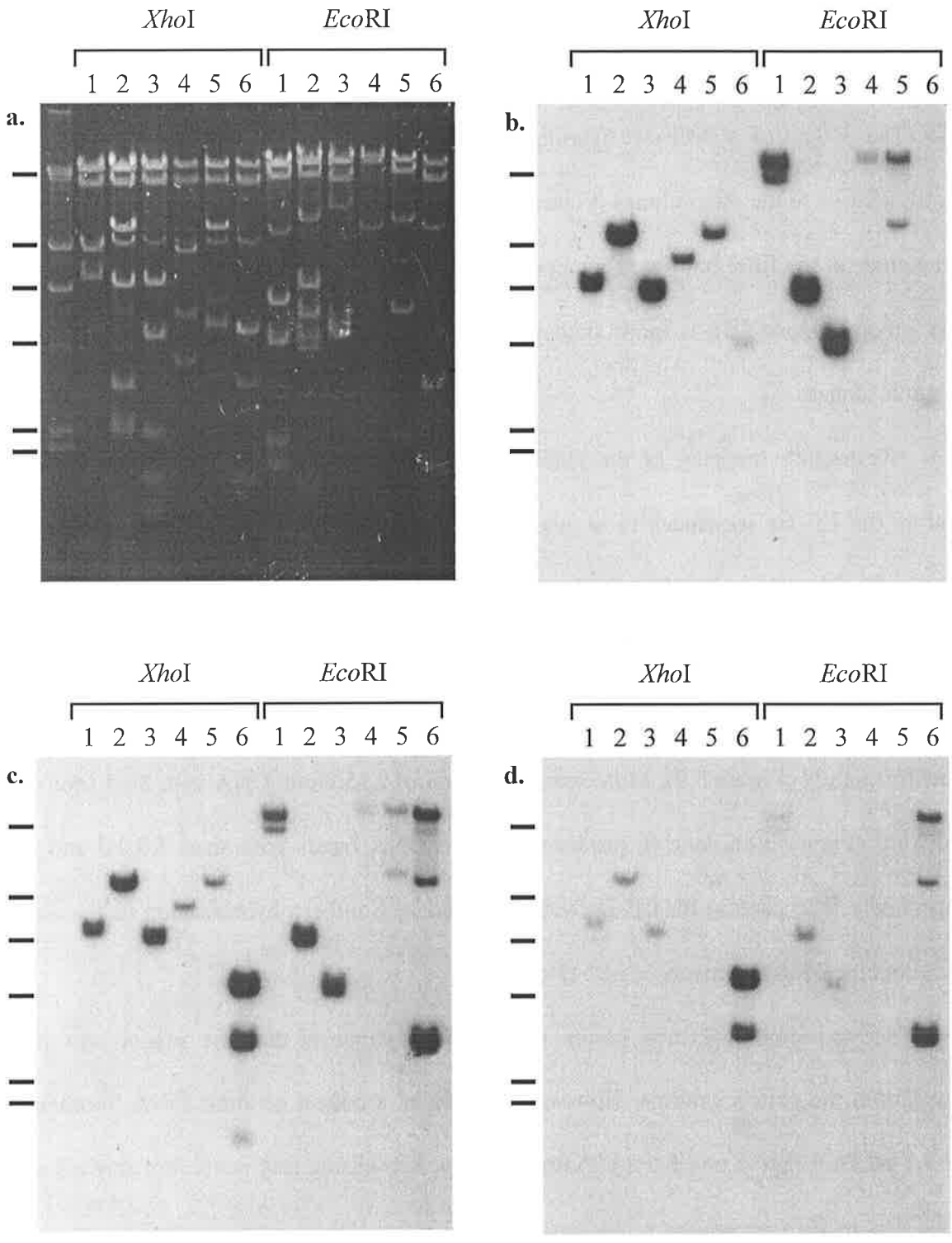
a. Agarose gel electrophoresis of DNA fragments resulting from restriction of EF-1α genomic clones with *Xho*I and *Eco*RI as indicated. The first lane contains molecular weight markers (Lambda DNA restricted with *Hin*dIII) with fragment sizes 23.1 kb, 9.4 kb, 6.6 kb, 4.4 kb, 2.3 kb and 2.0 kb. The remaining lanes each contain approximately 4 μg of recombinant phage DNA from four strongly-hybridising plaques (lanes 1-4) and two plaques which hybridised weakly to pGhEF-1α (lanes 5-6):

1. λL2;
2. λL3;
3. λL4;
4. λL5;
5. λS3;
6. λS5.

b. Southern blot of the gel depicted in a., hybridised with the insert from cDNA clone pGhEF-1α. Lanes and molecular weight markers are labelled as above.

c. Southern blot of the gel depicted in a., hybridised with the 188 bp PCR product resulting from EFP1/T7 amplification of cDNA clone pGhEF-1α. Exposure time was approximately 20 times longer than in part a. and lanes and molecular weight markers are labelled as above.

d. Southern blot from part c., after high-stringency washing of the filter. Exposure time was identical to that in part c. and lanes and molecular weight markers are labelled as above.


a.
XhoI
EcoRI
1 2 3 4 5 6 1 2 3 4 5 6
b.
XhoI
EcoRI
1 2 3 4 5 6 1 2 3 4 5 6
c.
XhoI
EcoRI
1 2 3 4 5 6 1 2 3 4 5 6
d.
XhoI
EcoRI
1 2 3 4 5 6 1 2 3 4 5 6

corresponded to EF-1α sequences and an 88 bp fragment which corresponded to pBluescript® SK(-) polylinker (not shown).

Hybridisation of the genomic clone Southern with the 188 bp fragment produced a pattern similar to that obtained with use of the whole cDNA probe (Figure 3.8c). The hybridisation pattern was identical to that depicted in Figure 3.8b except for genomic clone λS5. The 3' end of pGhEF-1α hybridised more strongly to λS5 than did the whole cDNA insert, relative to the other clones (Figure 3.8b, c). A wash at high stringency removed many of the counts on the filter but the probe bound to λS5 was retained (Figure 3.8d) which suggests that genomic clone λS5 is more similar to cDNA clone pGhEF-1α than the remaining five genomic clones.

Restriction mapping of the putative positive clone λS5 was undertaken in order to confine the EF-1α sequences to a subclonable DNA fragment. Of five restriction enzymes tested, *Eco*RI, *Xho*I, *Bam*HI, *Sac*I and *Xba*I, three (*Eco*RI, *Xho*I and *Sac*I) contained recognition sites within the λS5 insert sequence. Attempts to map the *Eco*RI sites were unsuccessful, but the estimated positions of the *Xho*I and *Sac*I sites suggest that the clone has a repetitive nature (Figure 3.9). Moreover, restriction of λS5 clone DNA with *Sac*I (not shown) and *Xho*I (Figure 3.8a, lane 6) produced doublet DNA bands with sizes 3.5 kb and 3.4 kb respectively. The gene(s) for EF-1α were localised by Southern hybridisation to the *Sac*I and *Xho*I doublet DNA fragments of λS5 (Figure 3.9).

The suspected repetitive nature of the clone suggests that the repeat unit may be prevalent in the cotton genome. However, probing of a cotton genomic DNA Southern with the 3.5 kb *Xho*I repeat unit from λS5 revealed few hybridising fragments (not shown) and the probe was ineffective against a Northern blot of cotton RNAs. Extensive efforts to subclone the fragments of λS5 which hybridised the cDNA clone by both gel purification of DNA fragments and purification of the whole clone were unsuccessful. Attempts to subclone the

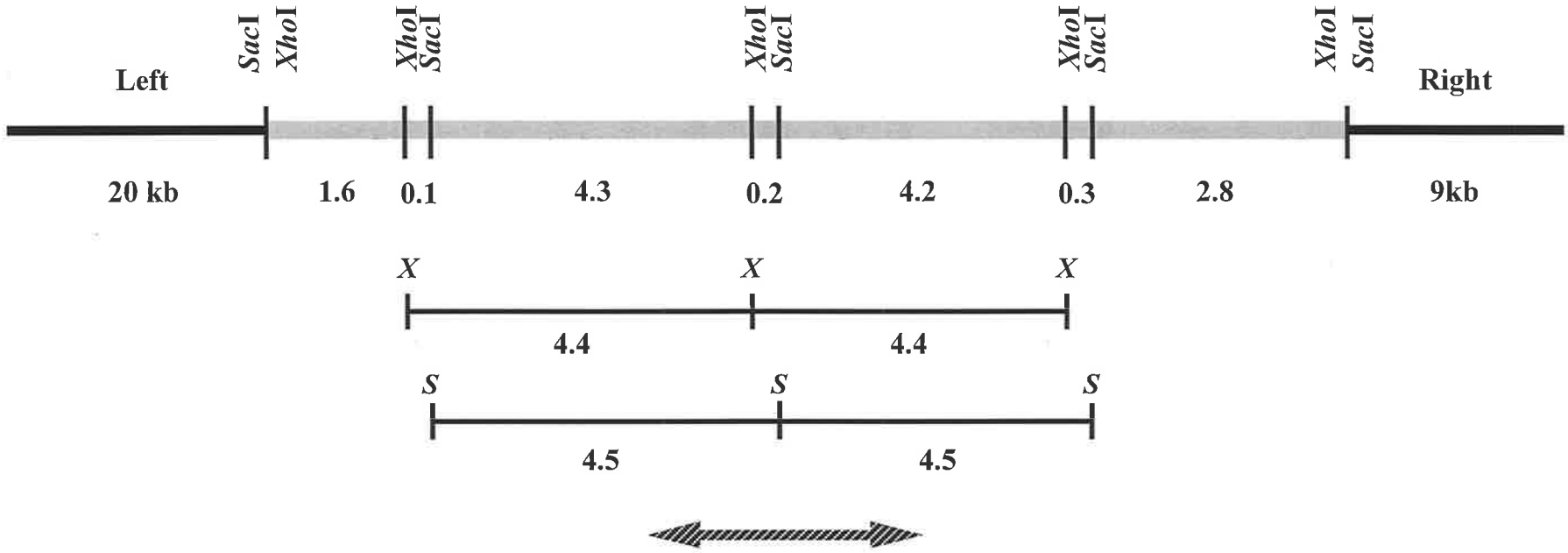


Figure 3.9: Partial restriction map of an elongation factor 1 α genomic clone from cotton

Restriction map of genomic clone λS5, showing the positions of the *Xho*I and *Sac*I sites, with fragment sizes given in kb and the region(s) hybridised by pGhEF-1α indicated by an arrow.

*Eco*RI fragments of interest from genomic clones λL4 and λS5 were also unsuccessful and further analysis of the cotton fibre EF-1α was not pursued.

3.7 Summary and Conclusions

Lack of a repeatable method for RNA extraction prevented purification of poly(A)$^+$ RNAs and resulted in a cotton fibre cDNA library which was rich in ribosomal and organellar transcripts. The predominance of cDNAs derived from poly(A)$^-$ RNAs clearly made the task of isolating fibre-specific cDNAs more labour-intensive. In addition, pooling of RNAs from fibres of different ages may dilute fibre-specific transcripts which might be expressed at a particular stage of fibre development. The isolation of fibre-specific cDNAs was unsuccessful in this instance.

Analysis of "cold" cDNA clones resulted in the identification of an interesting cDNA, pGhEF-1α, which most likely encodes a member of the superfamily of EF-1α proteins. EF-1α is universally conserved and is abundant in tissues which are active in protein synthesis. Given the abundance of EF-1α protein and mRNA in other systems, it is perplexing that the cotton fibre homologue was isolated as a "cold" cDNA clone. Southern analysis suggests that EF-1α is encoded by a family of related genes in the cotton genome, six of which were isolated as genomic clones. One of the clones, λS5, may encode pGhEF-1α, but this was unconfirmed experimentally.

CHAPTER 4

Abundant mRNAs Specific to the Developing Cotton Fibre

4.1 INTRODUCTION

An initial attempt to clone fibre-specific mRNAs was unsuccessful (Chapter 3), largely due to a lack of starting material, namely RNA. The development of a novel method for RNA extraction (Wan and Wilkins, 1994) allowed isolation of cellular RNA in high and reproducible yields from all cotton tissues including fibres. The following Chapter describes the characterisation of several fibre-specific cDNA clones isolated from cotton fibres.

4.2 ISOLATION OF PUTATIVE FIBRE-SPECIFIC cDNA CLONES

Differential screening of a cotton fibre cDNA library was undertaken using poly(A)$^+$ RNA from fibres of a single age. Cotton fibres aged 13 DPA precede the transition between the elongation and secondary wall deposition phases of fibre development (Basra and Malik, 1984) and analysis of 13 DPA fibres provides the possibility that cDNAs associated with either or both of the two processes will be isolated. Differences in the molecular profiles of cotton fibres during the stages of elongation (primary wall synthesis) and secondary wall synthesis are substantial (Alexander *et al.*, 1983; Timpa and Triplett, 1993), but results from experiments by Graves and Stewart (1988a) suggest that proteins involved in secondary wall synthesis are present as early as 10 DPA. In addition, 13 DPA fibres are of sufficient length to be harvested readily and lack the thick secondary walls which limit RNA extraction from older cotton fibres.

Monitoring of cDNA synthesis by incorporation of [α-^{32}P]dCTP yielded a single band of cDNA superimposed over a smear of products (not shown), similar to the result obtained previously (Figure 3.1a). The band, which is formed as a by-product of the enzymology used to perform the Gubler and Hoffman (1983) method of cDNA synthesis, also occurred with the use of control mRNA, whilst the negative control yielded negligible amounts of cDNA (not shown). Since the band was not apparent in the absence of RNA template, its presence is indicative of a cDNA product.

Cloning of the cDNA product from 13 DPA fibre RNA resulted in a library consisting of 2.3 x 10^4 recombinant clones, which was differentially screened with labelled cDNAs from leaf and 13 DPA fibre. Of the recombinant clones, 85% hybridised strongly to one or both probes whilst 15% did not hybridise to either of the two probes, suggesting that these "cold" plaques contained clones of rare mRNAs. Probing of a fraction of the library with the cotton rDNA clone pGhR1 did not identify any positives and copies of ribosomal RNA in the library were therefore assumed to be negligible.

Use of the conventional technique of +/- hybridisation (Sargent, 1987) resulted in the identification of 24 putative fibre-specific cDNA clones (Figure 4.1). The hybridisation pattern of each clone was more obvious after a secondary round of plaque purification, and three examples are shown in Figure 4.2. PCR across each insert revealed the relatively short length of the inserted cDNAs, which ranged in size from approximately 0.2 kb to 2 kb (Figure 4.3a). Cross-hybridisation eliminated duplicate clones and reduced the fibre-specific clone population to six different sequence types (Figure 4.3b). The pFS1 insert hybridised five other clones in addition to itself and clones homologous to pFS6 were the most abundant, comprising half of the 24 fibre-specific cDNAs. Sequences homologous to pFS17 occurred in two ~~longer~~ other clones, while pFS3, pFS18 and pFS20 were unique in the cDNA clone population. The clones named in Figure 4.3b are representative of each clone group and will be referred to throughout this thesis.

4.3 Expression Patterns of Putative Fibre-specific Genes

To confirm the spatial and investigate the temporal expression patterns of the cDNA clones, RNA blot analysis was carried out using RNA from various cotton tissues and from fibres at different developmental stages. Five of the six clones (pFS1, pFS3, pFS6, pFS17 and pFS18) showed highly preferential hybridisation to fibre mRNA (Figure 4.4), hybridising strongly to transcripts in fibre RNA but not to RNAs from leaf, whole flower or seedling

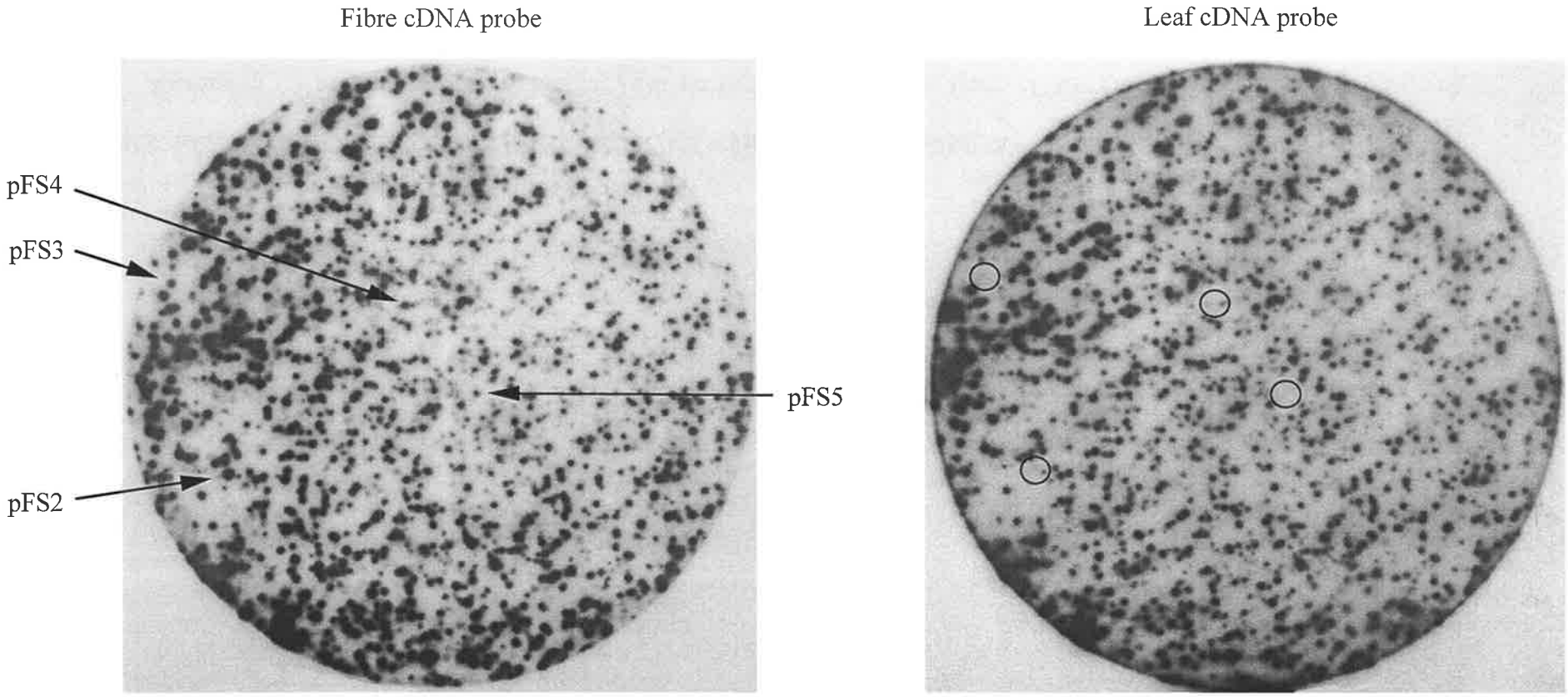


Figure 4.1: Isolation of fibre-specific cDNA clones

A sample of the 13 DPA fibre cDNA library, differentially screened with 13 DPA fibre cDNA (left) and leaf cDNA (right). Fibre-specific clones pFS2, pFS3, pFS4 and pFS5 are arrowed on the fibre autoradiograph and the corresponding areas on the leaf autoradiograph are circled.

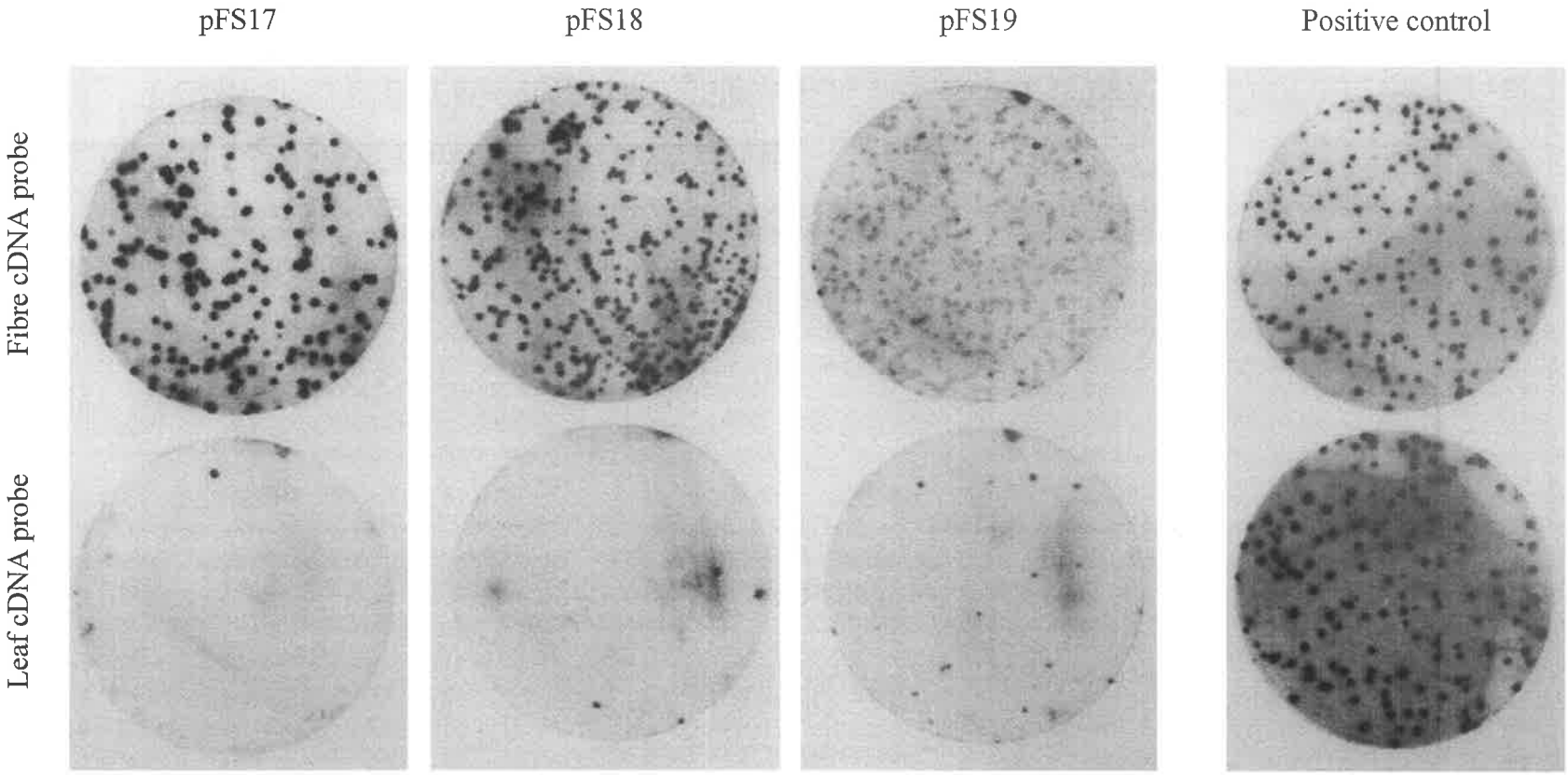


Figure 4.2: Secondary screening of putative fibre-specific cDNA clones

Examples of secondary screen plaque lifts probed with 13 DPA fibre cDNA (top) and leaf cDNA (below). Autoradiographs of pFS17, pFS18 and pFS19 are shown, together with a positive control to the right.

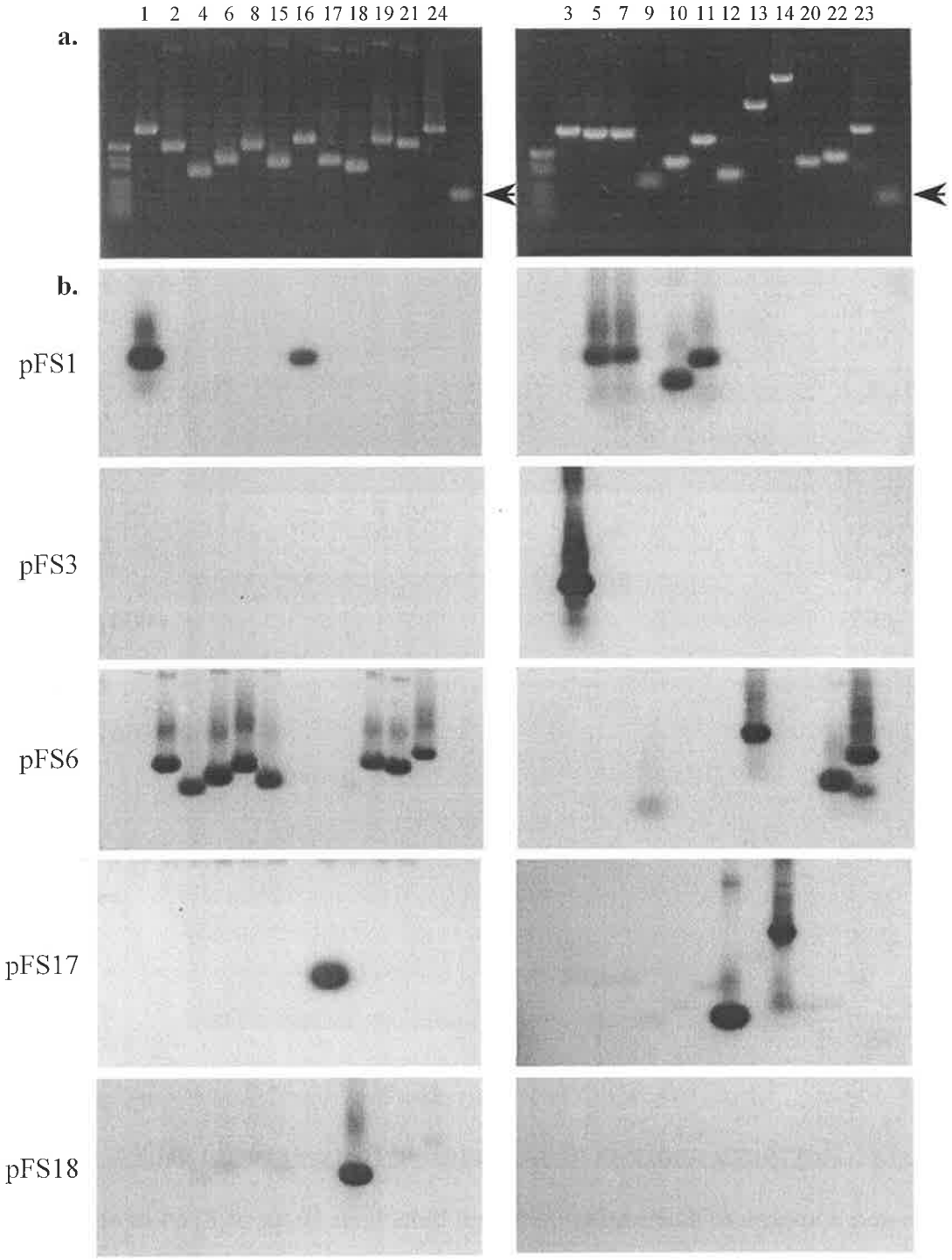


Figure 4.3: Cross-hybridisation of putative fibre-specific cDNA clones

a. Agarose gel showing PCR products resulting from amplification across the insert of each fibre-specific cDNA clone. Molecular weight markers (pUC19 DNA restricted with *Hpa*II) are shown on the left of each gel and the fragment sizes are 501 bp, 489 bp, 404 bp, 331 bp, 242 bp, 190 bp, 147 bp and 111 bp. Numbers at the top of each lane refer to clone number and arrows indicate the position of the 164 bp pBluescriptSK(-) control PCR product.

b. Hybridisation of a Southern blot of the above gel with various clones from the fibre-specific clone population, indicated at left.

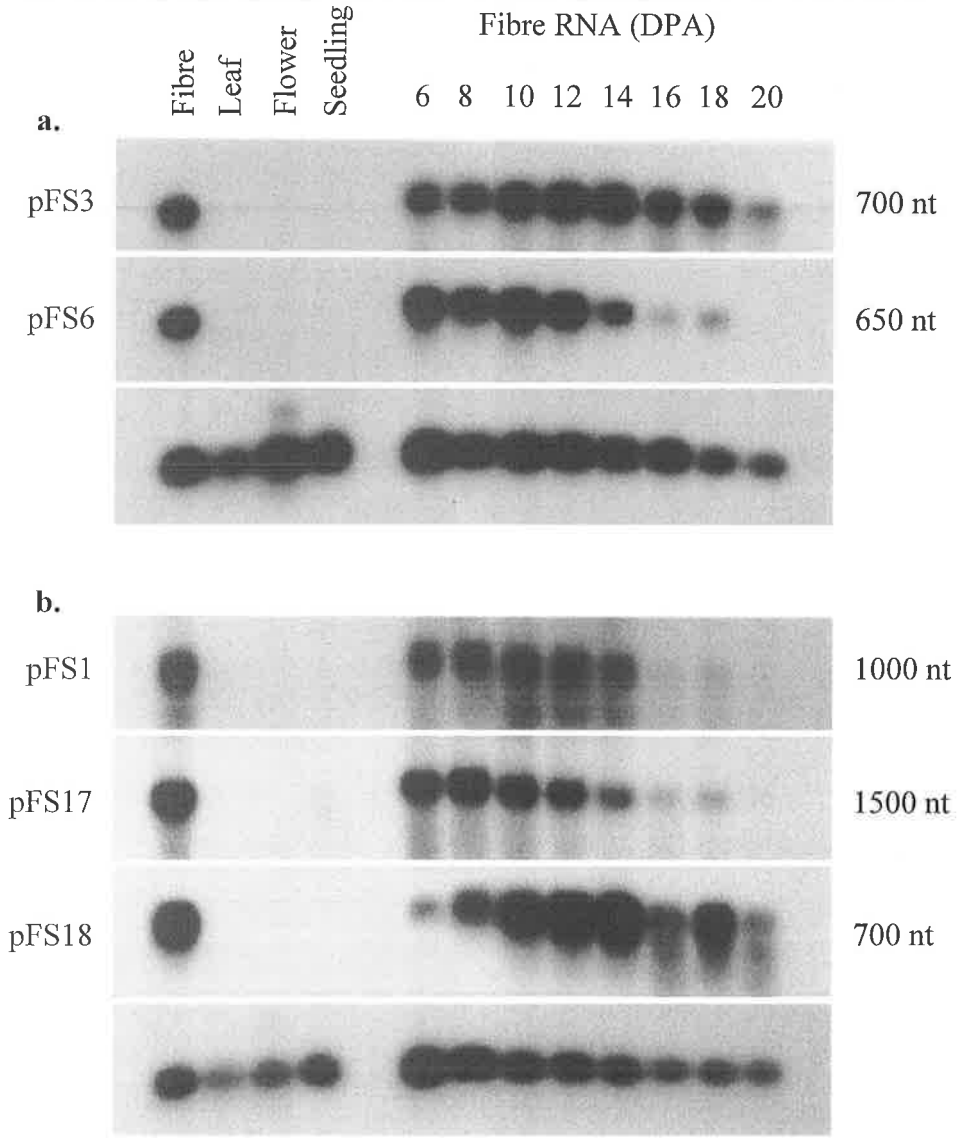


Figure 4.4: Northern analysis of six putative fibre-specific cDNA clones

a. and b. represent independent Northern blots with 10 μg of RNA in each lane, sequentially hybridised with the specific probes indicated. Total RNA loadings in each track were monitored by hybridisation with cDNA clone pFS20 and shown in the lower panel of each blot. Each clone detected a single transcript, the approximate sizes of which are indicated on the right.

The pFS17 hybridisation pattern was obtained using a 5'-RACE clone, pFS17-C (see Section 4.5.3b). The pFS18 hybridisation pattern resulted from probing with a longer homologue, identified by rescreening the cDNA library with pFS18.

tissue. On longer exposure, leaf, flower and seedling tissues showed weak hybridisation signals to each fibre-specific cDNA probe (not shown), indicating that transcription is not entirely restricted to the fibre or that products of genes with similar sequences cross hybridise to the cDNA probes. In most cases, genes which are expressed predominantly in one plant organ are also active in other organs (Edwards and Coruzzi, 1990).

The temporal pattern of transcript accumulation varied between clones, as shown by hybridisation of cDNA inserts to RNA from fibres of varying ages (Figure 4.4, lanes 6-20). The timespan 6-20 DPA encompasses the transition from elongation of the fibre cell to secondary wall deposition. Secondary wall synthesis begins at approximately 16 DPA, and there is an overlap of several days between the two phases of fibre development (Section 1.3.4). As expected, all transcripts were abundant in 13 DPA fibre tissue, which was the source of the mRNA for the differential screen. Transcripts corresponding to pFS1, pFS6 and pFS17 increased in concentration as fibre development proceeded through primary wall synthesis, or elongation (Figure 4.4, lanes 6-16), and decreased at the onset of secondary wall synthesis. The synthesis and accumulation of a large number of cell wall-associated proteins are increased during elongation, such as those observed in carrot suspension cells (McCann and Roberts, 1994). This suggests a role for the proteins which correspond to pFS1, pFS6 and pFS17 in primary wall synthesis or elongation of the cotton fibre.

Transcripts of pFS3 were present during both the primary and secondary cell wall synthesis stages (Figure 4.4, lanes 6-20), suggestive of a continuous role throughout fibre cell development. The abundance of pFS18 mRNA appeared to peak at 12 to 14 DPA (Figure 4.4, lanes 12-14), which precedes the transition between the elongation (primary wall synthesis) and secondary wall deposition phases of fibre growth. The protein(s) encoded by pFS18 may therefore be involved in primary wall synthesis, secondary wall deposition or in the biochemical and physiological changes which are associated with the transition between the two phases (Section 1.3.4). While it is unknown whether these mRNAs are translated, differing temporal

expression patterns probably indicate a role for transcriptional control of gene expression in cotton fibre cells. Differential stability of mRNAs may also account for the distortions.

One clone, pFS20, was unique in the clone collection and did not represent a fibre-specific transcript. Rather, pFS20 transcripts were present equally in fibre, leaf, flower and seedling tissues, producing constitutive signals comparable with those obtained for the cotton rDNA clone pGhR1 (not shown). Hybridisation with this probe provides a useful positive control, reflecting RNA loading (Figure 4.4). Subsequent sequencing of pFS20 yielded 210 bp of sequence which lacked a poly(A) tail and showed at least 90% nucleotide identity to mitochondrial 26S rRNA genes from a variety of plant species including maize (Dale *et al.*, 1984), wheat (Falconet *et al.*, 1988) and potato (Binder *et al.*, 1994). Although pFS20 was clearly fibre-specific in the differential screen, there may have been an error in the selection of the original plaque or in the selection of a positive plaque from the secondary screen plate of pFS20. Secondary screening of the second batch of fibre-specific cDNAs (depicted on the right in Figure 4.3b and including pFS20) was inefficient with respect to the first twelve clones and hybridisation of labelled leaf cDNA to the positive control plaques was weak, making the latter explanation probable. The majority of pFS20 hybridisation occurred to a transcript of approximately 300 nt in size (Figure 4.4) but several larger RNA fragments also showed homology to pFS20 (not shown). These rRNAs, together with the 300 nt rRNA species, result from degradation of the mitochondrial rRNA into specific products (Leaver and Ingle, 1971). pFS20 is therefore probably an inappropriate control for RNA loading and integrity, but probing of Northerns with the 5S rRNA from rye, pF15 (Table 2.2), was ineffective.

The estimated sizes of the transcripts specified by the five fibre-specific clones ranged between 600 nt and 1500 nt, as indicated in Figure 4.4. Based on a comparison of cDNA lengths of clones in each group (Figure 4.3a) and transcript size from RNA blots (Figure 4.4), it was concluded that full-length clones were not obtained for pFS1, pFS3, pFS17 and pFS18. However, there were several clones in the pFS6 group which contained inserts of the size

expected for a full-length transcript. Two such clones, pFS23 and pFS24, contained inserts sized in the order of 650 bp (Figure 4.3a), and clone pFS13 contained a larger insert, possibly containing a pFS6 cDNA together with one or more independent cDNAs.

4.4 Relative Abundance of Fibre-specific Transcripts

A limitation of the differential screening technique is that it may only be used in detection of abundant mRNAs which represent at least 0.1% of the total mRNA population (Sargent, 1987). The relative abundance of each fibre-specific transcript was determined by using isolated cDNA inserts as probes to a larger sample of the 13 DPA fibre cDNA library, consisting of 3.7×10^4 recombinants. The results are summarised in Table 4.1.

Table 4.1: Relative abundance of fibre-specific cDNA clones in 13 DPA fibres

Representative clone	Number of clones in group	Number of positive cDNA clones (Total 3.7×10^4)	Transcript abundance (% of cDNA)
pFS1	6	156	0.47%
pFS3	1	53	0.16%
pFS6	12	215	0.65%
pFS17	3	75	0.23%
pFS18	1	58	0.18%

The five families of fibre-specific transcript make up a substantial proportion of the mRNA population in 13 DPA fibre tissue, and values greater than 0.1% were obtained in all cases. This conclusion is based on the assumption that each poly(A)$^+$ mRNA is equally clonable and that each clone contains a single cDNA insert. The likelihood of multiple inserts was apparent in at least one fibre-specific clone, pFS13, indicating that the mRNA abundance values may be an over-estimation and that fibre-specific transcripts may be rarer than shown here. Relatively however, pFS6 transcripts were the most abundant fibre-specific mRNAs in 13

DPA fibres and comprise about 0.65% of the total mRNA population. Transcripts corresponding to pFS1, pFS17, pFS3 and pFS3 constitute rarer classes of mRNA, comprising 0.47%, 0.23%, 0.16% and 0.18% of the total mRNA respectively. Transcript abundance correlates positively with the number of clones in each group (Table 4.1, columns 2 and 4).

The apparent discrepancy between the isolation of only 24 original clones and large numbers of complementary plaques in the same cDNA library (Table 4.1, column 3) arose because only clearly identified plaques were picked. In addition, many differentially hybridising plaques were overlooked in the original screen. Examination of the autoradiographs of the original differential screen showed that the majority of clones identified by rescreening the library were fibre-specific, but some plaques showed a small amount of hybridisation to the leaf cDNA probe. It may be that the cDNA clones are detecting multitranscripts or transcripts from a multigene family, some members of which may not be differentially expressed. This may also contribute to an overestimation of fibre-specific transcript abundance. As expected, each clone detected plaques corresponding to the original clones from its own group of fibre-specific cDNAs, but cross-hybridisation between clones of different groups was not observed.

4.5 Sequencing and Southern Analysis of Fibre-specific cDNA clones

Sequence information for cDNA clones corresponding to each of the five differentially expressed mRNAs was used to determine potential functions of the encoded proteins. In the cases of pFS1 and pFS17, the longest clone in each group was sequenced, and the longest single-insert clone in each group was chosen for sequence analysis. The results of a detailed characterisation of the pFS6 clone group are presented in the following Chapter.

4.5.1 The pFS1 Clone Group

4.5.1a Sequence Analysis of pFS1 and its Homologues

Clone pFS5 is the longest clone of the pFS1 group (Figure 4.3) but was too small to

contain the full-length pFS1 transcript. The pFS5 insert was sequenced in its entirety, yielding a sequence of 509 bp which lacked a poly(A) tail (Figure 4.5). The reasons for a lack of poly(A) tail have been discussed previously (Section 3.3). The sequence of pFS5 showed 96% nucleotide identity with the corresponding region of a previously characterised fibre-specific cDNA, E6 (John and Crow, 1992; Section 1.6.2). The differences included eight single nucleotide substitutions and three deletions from the pFS5 sequence as compared to the E6 cDNA sequence (Figure 4.6). The deletions of nine, three and one bp occurred at positions 523-531, 931-933 and 900 respectively in the E6 sequence. The 91 residue amino acid sequence derived from clone pFS5 (Figure 4.5) is 91% identical to the 94 carboxy-terminal amino acids of the E6 protein (John and Crow, 1992; Figure 4.6). A significant difference is the in-frame deletion of nine nucleotides from pFS5 compared with E6, which results in the absence of three amino acids from the conceptual translation product of pFS5. In addition, the protein sequences contain five amino acid substitutions, of which two are conservative changes. These sequence differences may be ascribable to allelic differences between *G. hirsutum* L., cv. Coker 312 and *G. hirsutum* L., cv. Siokra 1-2 or to differences between homologues from the *A* and *D* subgenomes of *G. hirsutum*.

The function of the E6 protein is unknown but it has been postulated to have a unique structural or enzymatic role in cotton fibre primary cell wall deposition. This is consistent with results from Southern analysis, which show conservation of an E6-homologous gene in other fibre-producing plants such as kapok (John and Crow, 1992). A direct role in cellulose synthesis can be ruled out because the maximum rate of cellulose deposition in cotton fibre occurs between 26 and 28 DPA, when E6 protein concentration is very low (John and Crow, 1992).

Predominance of pFS1-related transcripts early in fibre development (Figure 4.4a) contrasts with the expression pattern of E6, which showed maximum mRNA levels in fibres aged 15-22 DPA (John and Crow, 1992). However, early accumulation of the pFS1 transcript

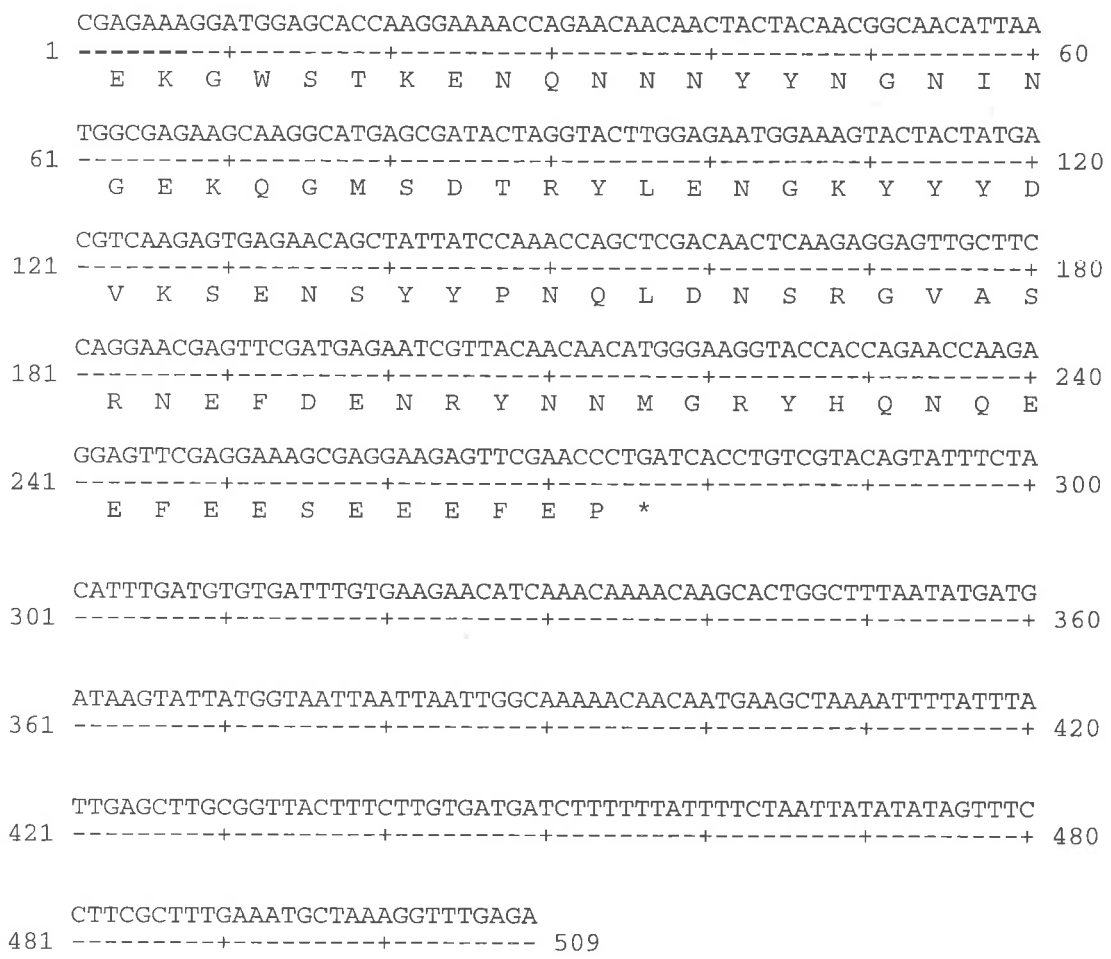


Figure 4.5: Nucleotide sequence and conceptual translation of clone pFS5

Nucleotide sequence of fibre-specific cDNA clone pFS5, showing conceptual translation from reading frame #2 in single letter code below the cDNA sequence. The putative stop codon, TGA, is indicated by an asterisk. Polyadenylation signal(s) were not apparent in the sequence of pFS5, which lacked a poly(A) tail.

correlates with the immunodetection of the E6 protein product in Western analysis (John and Crow, 1992). Maximal E6 protein accumulation was observed to occur between 5 and 15 DPA and the discrepancy between this and the maximal presence of the corresponding E6 transcript was attributed to differential stability of the protein. The E6 protein contains several sequence motifs that indicate potential post-translational modifications (John and Crow, 1992). Environmental factors ~~or varietal differences~~ are perhaps more likely to be responsible for the observed difference in pFS5 and E6 expression patterns than varietal influences. Cotton plants of cv. Coker 312 were grown in the greenhouse (John and Crow, 1992), whereas in this experiment, Siokra 1-2 plants were grown under controlled conditions in plant growth cabinets. The growth and development of cotton fibres are greatly influenced by environment and environment-genotype interactions (O'Kelley and Carr, 1953; Gipson and Joham, 1969; Gipson and Ray, 1969) and the effects of factors such as temperature, soil moisture and light are well-documented (Section 1.1.4).

John and Crow (1992) suggest that E6 is the most abundant "fibre-specific" transcript in cotton fibres. Since our results suggest that pFS6-related transcripts are the most abundant fibre-specific mRNA in fibre tissue (at least in fibres aged 13 DPA), this may constitute further evidence for environmental and varietal differences between *G. hirsutum* L., cv. Coker 312 and *G. hirsutum* L., cv. Siokra 1-2.

4.5.1b Identification of pFS1 Genomic Clones

Screening of a Siokra 1-2 partial genomic library consisting of 1.6×10^5 clones revealed 28 positive plaques, which varied in their size and intensity of hybridisation to the pFS1 probe (not shown). The promoter of the pFS1 homologue, E6, has been extensively characterised and shown to direct the expression of a reporter gene in a tissue-specific and developmentally-regulated fashion in transgenic cotton plants (John and Crow, 1992). Consequently, further analysis of pFS1 genomic clones was not pursued.

Figure 4.6: Sequence comparison between cDNA clones pFS5 and E6

Alignment of the pFS5 sequence with the full-length E6 cDNA sequence, with nucleotide differences indicated by asterisks below the sequence alignment. The sequence of pFS5 terminates at nucleotide position 986 in the E6 sequence. The conceptual translation of the E6 cDNA in single letter code above the cDNA sequence and the putative polyadenylation signal in the sequence of E6 is double underlined. The putative stop codon TGA is indicated by a (•).

```
                                                  M  G  S  S  P  K  L  F  S
E6      ACACACACAAGTAAAGCATTAGCAACCATAGCCATGGGTTCCTCACCAAAACTCTTCTCT    60
pFS5    ------------------------------------------------------------

         M  S  I  L  F  L  F  A  P  F  S  M  Q  I  H  A  R  E  Y  F
E6      ATGTCTATCCTCTTCCTTTTTGCCCCCTTCTCCATGCAAATCCATGCTAGAGAGTACTTC   120
pFS5    ------------------------------------------------------------

         S  K  F  P  R  V  N  I  N  E  K  E  T  T  T  R  E  Q  K  H
E6      AGCAAATTCCCAAGAGTTAACATCAATGAGAAAGAGACAACAACCAGAGAGCAAAAGCAC   180
pFS5    ------------------------------------------------------------

         E  T  F  V  P  Q  T  T  Q  K  P  E  E  Q  E  P  R  F  I  P
E6      GAGACCTTCGTTCCCCAGACCACCCAAAAGCCAGAAGAACAAGAGCCAAGGTTCATTCCT   240
pFS5    ------------------------------------------------------------

         E  T  Q  N  G  Y  G  L  Y  G  H  E  S  G  S  S  R  P  S  F
E6      GAAACCCAAAATGGTTATGGCCTTTACGGCCACGAGTCAGGCTCAAGCCGGCCCAGTTTC   300
pFS5    ------------------------------------------------------------

         T  T  K  E  T  Y  E  P  Y  V  T  P  V  I  F  H  P  D  E  P
E6      ACCACCAAAGAAACCTATGAACCCTATGTCACCCCTGTTATATTCCACCCTGATGAGCCC   360
pFS5    ------------------------------------------------------------

         Y  N  S  I  P  E  S  S  N  N  K  D  T  Y  Y  Y  N  K  N  A
E6      TATAACAGCATCCCCGAATCCTCCAACAATAAAGACACTTACTACTACAACAAGAATGCC   420
pFS5    ------------------------------------------------------------

         Y  E  S  T  K  Q  Q  N  L  G  E  A  I  F  T  E  K  G  W  S
E6      TACGAGTCCACTAAGCAGCAAAACTTGGGCGAGGCCATTTTCACCGAGAAAGGATGGAGC   480
pFS5    ----------------------------------------CGAGAAAGGATGGAGC    16

         T  K  E  N  Q  N  N  N  Y  Y  N  G  N  N  G  Y  N  N  G  E
E6      ACCAAGGAAAACCAGAACAACAACTACTACAACGGCAACAATGGTTACAACAATGGCGAG   540
pFS5    ACCAAGGAAAACCAGAACAACAACTACTACAACGGCAACATT---------AATGGCGAG    67
                                                 * *********

         K  Q  G  M  S  D  T  R  Y  L  E  N  G  K  Y  Y  Y  D  V  K
E6      AAGCAAGGCATGAGCGATACTAGGTACTTGGAGAATGGAAAGTACTACTATGACGTCAAG   600
pFS5    AAGCAAGGCATGAGCGATACTAGGTACTTGGAGAATGGAAAGTACTACTATGACGTCAAG   127

         S  E  N  N  Y  Y  P  N  R  F  D  N  S  R  G  V  A  S  R  N
E6      AGTGAGAACAACTATTATCCAAACCGGTTCGACAACTCAAGAGGAGTTGCTTCGAGGAAC   660
pFS5    AGTGAGAACAGCTATTATCCAAACCAGCTCGACAACTCAAGAGGAGTTGCTTCCAGGAAC   187
                  *                 * *                          *

         E  F  N  E  N  R  Y  N  N  M  G  R  Y  H  Q  N  Q  E  E  F
E6      GAGTTCAATGAGAATCGTTACAACAACATGGGAAGGTACCACCAGAACCAAGAGGAGTTC   720
pFS5    GAGTTCGATGAGAATCGTTACAACAACATGGGAAGGTACCACCAGAACCAAGAGGAGTTC   247
          *

         E  E  S  E  E  E  F  E  P  •
E6      GAGGAAAGCGAGGAAGAGTTCGAACCCTGATCACCTGTCGTACAGTATTTCTACATTTGA   780
pFS5    GAGGAAAGCGAGGAAGAGTTCGAACCCTGATCACCTGTCGTACAGTATTTCTACATTTGA   307

E6      TGTGTGATTTGTGAAGAACATCAAACAAAACAAGCACTGGCTTTAATATGATGATAAGTA   840
pFS5    TGTGTGATTTGTGAAGAACATCAAACAAAACAAGCACTGGCTTTAATATGATGATAAGTA   367

E6      TTATGGTAATTAATTAATTGGCAAAAACAACAATGAAGCTAAAATTTTATTTATTGAGCC   900
pFS5    TTATGGTAATTAATTAATTGGCAAAAACAACAATGAAGCTAAAATTTTATTTATTGAGC-   426
                                                                   *

E6      TTGCGGTTAATTTCTTGTGATGATCTTTTTTTTTATTTTCTAATTATATATAGTTTCCTT   960
pFS5    TTGCGGTTACTTTCTTGTGATGATCTTTTT---TATTTTCTAATTATATATAGTTTCCTT   483
              *                  ***

E6      TGCTTTGAAATGCTAAAGGTTTGAGAGAGTTATGTTCTTTTTCTCTTCCTCTTTCTTTTT  1020
pFS5    CGCTTTGAAATGCTAAAGGTTTGAGA----------------------------------   509

E6      TAACTTTATCAAACAATTTTTGAATAAAAATGTGAGTATATTGTAAC 1067
pFS5    ------------------------------------------
```

4.5.1c Southern analysis of fibre-specific clone pFS1

Southern analysis of genomic DNA using pFS1 as a probe produced two major hybridising fragments, together with several minor bands of hybridisation (Figure 4.7). The gene encoding pFS1 is probably present as one or two copies in the genome of *G. hirsutum* L., cv. Siokra 1-2, but there may be several related sequences in the genome, with weak homology to pFS1. The two genes may each be derived from one of the two ancestral diploid genomes which constitute the genome of *G. hirsutum* (Wendel, 1989). This result is in agreement with Southern results obtained by John and Crow (1992) for E6, in which one or two hybridising fragments were identified in restricted *Gossypium* genomic DNA.

4.5.2 Cotton Fibre-specific Clone pFS3

4.5.2a Initial Sequence Characterisation of Fibre-specific Clone pFS3

Clone pFS3 was unique in the fibre-specific cDNA clone population and did not hybridise to any other inserts (Figure 4.3b). Sequencing revealed an insert of 539 bp, including a poly(A) tail of 21 adenine residues (Figure 4.8). The transcript size of pFS3 was estimated from a Northern blot at approximately 700 nt (Figure 4.4), so despite the identification of two putative AUG initiation codons near the 5' terminus of the sequence (Figure 4.8), pFS3 was not expected to contain the full-length cDNA. Two putative polyadenylation signals (5'-AATAA'-3') were identified at the 3' end of the sequence, positioned 26 bp and 23 bp upstream from the site of polyadenylation. Screening of the nucleotide databases (GenBank and EMBL, September 1994) with the pFS3 cDNA sequence revealed no significant homology to any known sequences.

4.5.2b Isolation of the Full-length pFS3 cDNA by 5'-RACE

In order to obtain the full-length pFS3 cDNA, 5'-RACE was carried out, using primers specific to the pFS3 mRNA sequence (Figure 4.8; Table 2.3b). An *Eco*RI site was incorporated into the 5' end of the PCR primer, P3.2, to allow cloning of the 5'-RACE PCR

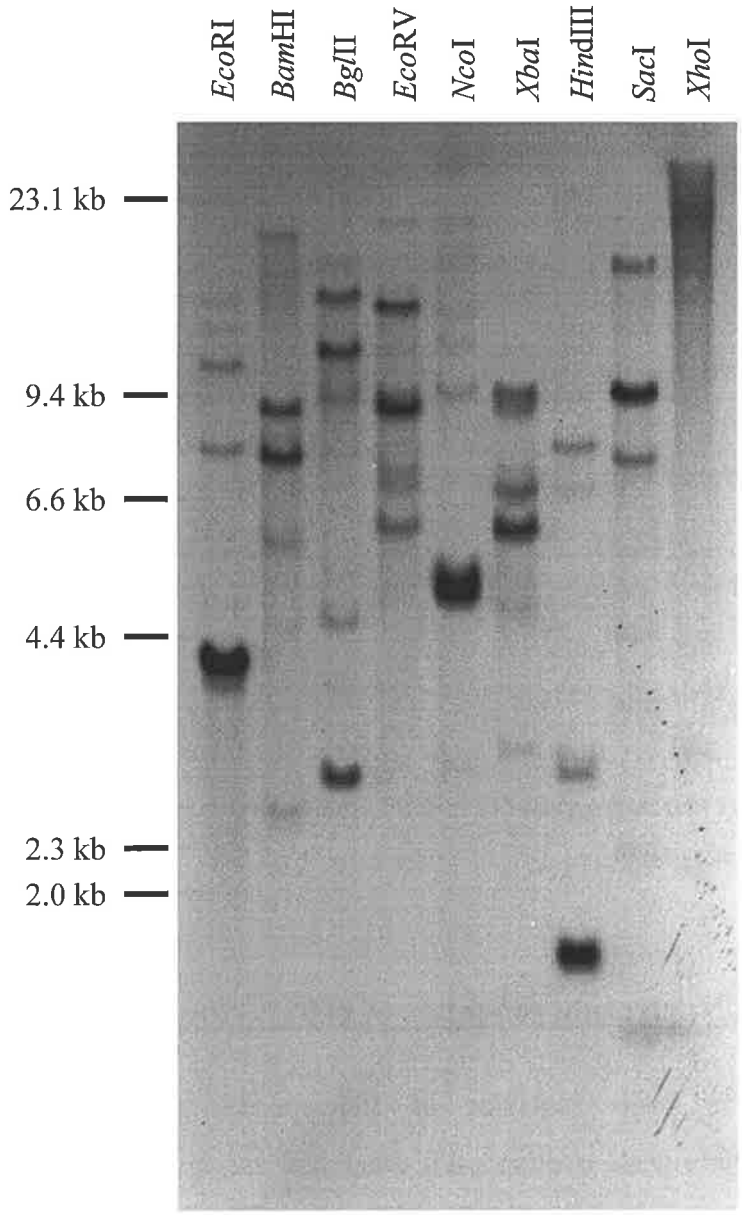


Figure 4.7: Southern analysis of fibre-specific cDNA clone pFS1

Genomic Southern blot probed with the pFS1 insert. Each lane was loaded with 10 µg of total genomic DNA from *G. hirsutum* L., cv. Siokra 1-2 cut with restriction enzymes, as indicated, that do not contain recognition sites within the pFS5 sequence. Numbers on the left represent the molecular weights of wild type Lambda DNA restricted with *Hin*dIII (kb).

Figure 4.8: Nucleotide sequence of cDNA clone pFS3

Complete insert sequence of fibre-specific cDNA clone pFS3. The putative polyadenylation signals overlap each other and are in bold and double underlined and the regions complementary to 5'-RACE cDNA primer (P3.1) and the 5'-RACE PCR primer (P3.2) are boxed. Possible ATG initiation codons at nucleotide positions 83 and 113 in the 5' end of the sequence are underlined.

product into a plasmid vector. The 5'-RACE experiment resulted in a single amplification product of approximately 350 bp in length (Figure 4.9a, lane 5) which was absent in the negative control (Figure 4.9a, lane 6). Since the PCR primer was designed 174 bp from the 5' terminus of the pFS3 cDNA sequence (Figure 4.8), the 350 bp product contained approximately 175 bp of additional 5' sequence. This would result in a full-length pFS3 sequence of approximately 770 bp, which is compatible with an estimated transcript size of 700 nt (Figure 4.4a). The generation of a 338 bp amplification product for the positive control (Figure 4.9a, lane 2) indicated that cDNA synthesis was successful but absence of such a band after the cDNA purification step (Figure 4.9a, lane 3) revealed that a loss of cDNA during purification or an error in the PCR reaction (such as lack of inclusion of a primer) had occurred. A few additional cycles of amplification may have resulted in an observable PCR product, since PCR using control DNA after purification and anchor ligation yielded the expected product of approximately 500 bp (Figure 4.9a, lane 4).

A Southern blot of the pFS3 5'-RACE gel was probed with the insert of pFS3, which showed strong hybridisation to the 5'-RACE PCR product and the pFS3 plasmid DNA (Figure 4.9b, lanes 5 and 7). Subsequent cloning of the 350 bp pFS3 5'-RACE PCR product resulted in the identification of four positive plasmid clones (Figure 4.9c, d), three of which, pFS3-26, pFS3-39 and pFS3-49, appeared identical and contained inserts of approximately 300 bp in size (Figure 4.9c, lanes 3-5). The fourth clone, pFS3-59, appeared to contain an insert with an internal *Eco*RI site (Figure 4.9c, lane 6) which together result in an inappropriate insert of approximately 500 bp. All insert fragments hybridised strongly to a pFS3 probe (not shown). PCR amplification across each insert using T3 and T7 primers resulted in a single, major PCR product in each case which were unexpectedly large, a result which is consistent with presence of multiple inserts in each clone. A product of estimated size 750 bp was generated for clones pFS3-26, pFS3-39 and pFS3-49 and a product of 650 bp was generated for clone pFS3-59 (not shown). Clearly then, each 5'-RACE clone contained two PCR products. The products

Figure 4.9: Cloning of the 5' end of cDNA clone pFS3

a. Agarose gel showing products resulting from PCR amplification of 5'-RACE products of pFS3. Lanes 2-6 are each loaded with a 10 µl sample of PCR product from a reaction volume of 50 µl and the lanes are:

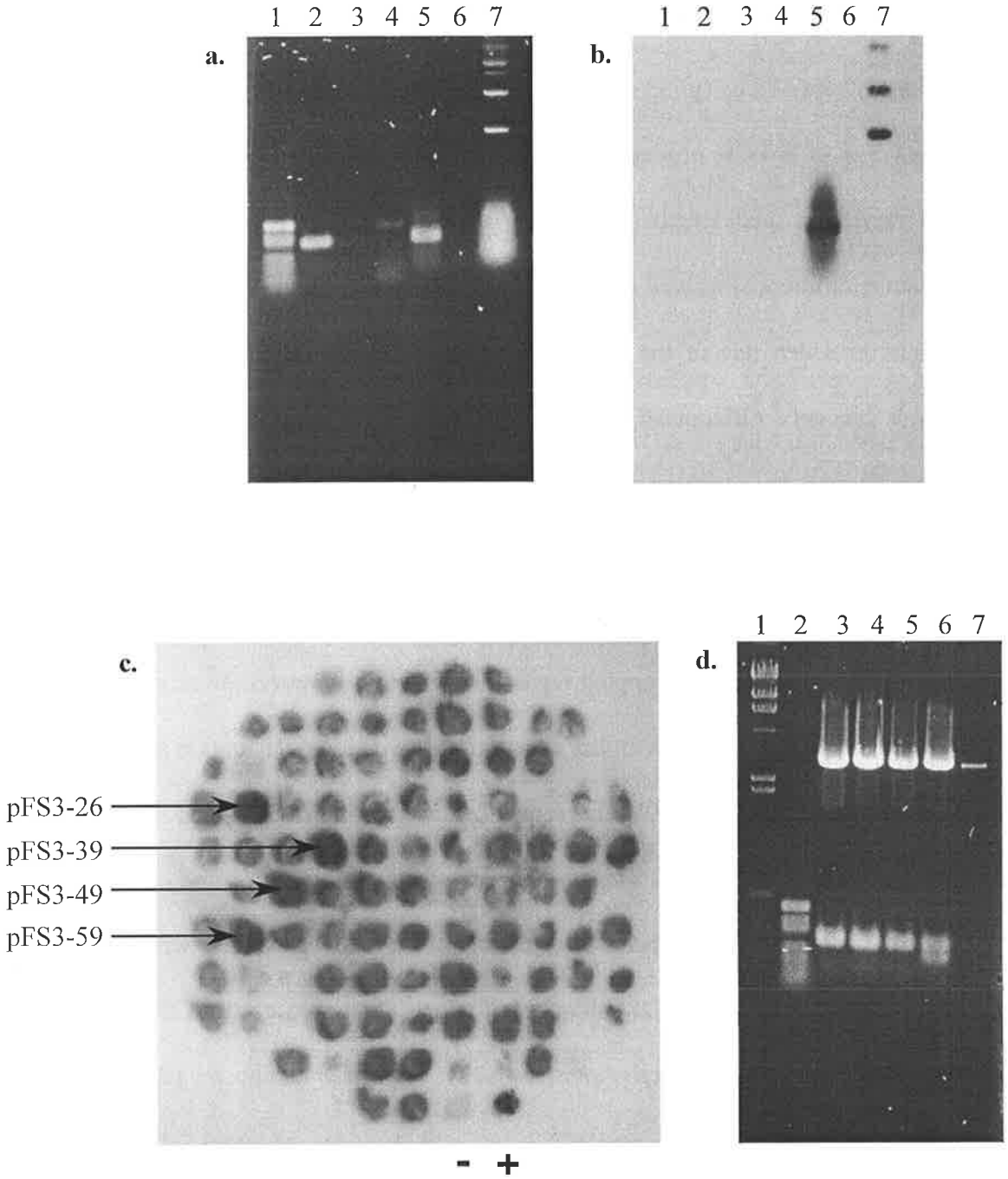
1. Molecular weight markers (pUC19 DNA restricted with *Hpa*II) ,with fragment sizes 501 bp, 489 bp, 404 bp 331 bp, 242 bp, 190 bp, 147 bp and 111 bp;
2. TFR control cDNA prior to purification, utilising gene-specific primers which amplify a product of 338 bp;
3. TFR control cDNA after the purification step, utilising gene-specific primers which amplify a product of 338 bp;
4. TFR control cDNA after both purification and anchor ligation, utilising one gene-specific primer together with the anchor primer. A product of 480 bp should be amplified;
5. 5'-RACE PCR of pFS3 cDNA, utilising gene-specific PCR primer P3.2 together with the anchor primer.
6. Negative control reaction, using the same primers as those used in the experiment, but without the addition of a cDNA template;
7. A 1 µg sample of uncut pFS3 plasmid DNA, included as a positive control for the subsequent Southern analysis.

b. Southern blot of the gel in part a., hybridised with the pFS3 cDNA insert. Lanes are numbered as in part a.

c. Colony hybridisation of putative pFS3 5-RACE clones, probed with the pFS3 insert. Positive clones pFS3-26, pFS3-39, pFS3-49 and pFS3-59 are indicated by arrows, and the positions of positive and negative control colonies are indicated by a (+) and (-) respectively.

d. Agarose gel showing the pFS3 5'-RACE clones identified above. Each lane is loaded with approximately 2 µg of plasmid DNA restricted with *Eco*RI and the lanes are:

1. Molecular weight markers (Lambda DNA restricted with *Hin*dIII) ,with fragment sizes 23.1 kb, 9.4 kb, 6.6 kb, 4.4 kb, 2.3 kb and 2.0 kb;
2. Molecular weight markers (pUC19 DNA restricted with *Hpa*II) ,with fragment sizes 501 bp, 489 bp, 404 bp 331 bp, 242 bp, 190 bp, 147 bp and 111 bp;
3. Clone pFS3-26;
4. Clone pFS3-39;
5. Clone pFS3-49;
6. Clone pFS3-59;
7. A 1 µg sample of pBluescript®SK(-), restricted with *Eco*RI.


a.
1 2 3 4 5 6 7
b.
1 2 3 4 5 6 7
c.
pFS3-26
pFS3-39
pFS3-49
pFS3-59
- +
d.
1 2 3 4 5 6 7

were identical in size in three of the clones (Figure 4.9d, lanes 3-5), but the two inserts differed in size in clone pFS3-59 (Figure 4.9d, lane 6).

4.5.2c Sequence Analysis of Four pFS3 5'-RACE Clones

In order to resolve the nature of the 5'-RACE PCR inserts and to obtain the sequence of the full-length pFS3 cDNA, the four clones were sequenced in their entirety. Insert sequence lengths ranged from 511 bp (pFS3-59), to 582 bp (pFS3-26), and all four clones contained two independent 5'-RACE PCR products, arranged in the same orientation with respect to each other and varying in their lengths. Alignment of the eight PCR product sequences revealed some sequence differences as well as some length variation (Figure 4.10). There are four base substitutions, in which one of the sequences differs from the others by a single nucleotide. These minor sequence differences may be attributed to *Taq* DNA polymerase errors during PCR or to reverse transcription errors during first-strand cDNA synthesis. All 5'-RACE products contained appropriate *Eco*RI sites and were flanked by a 5'-GTG-3' sequence at their 5' ends, which is the remnant of the anchor sequence, and the P3.2 sequence at their 3' end (Figure 4.10). As expected, overlapping regions were found between the sequences of the 5'-RACE clones and that of pFS3. With the exception of two single nucleotide substitutions, one in pFS3-49a and one in pFS3-59b, the sequences were identical to pFS3 in the region of overlap.

There is substantial variation in the length of the PCR products, with the pFS3-59b sequence providing only 11 bp of additional pFS3 clone sequence (a 5'-RACE PCR product of 266 bp). The remaining seven inserts can be divided into three categories (Figure 4.10). The first of these contains pFS3-39a and pFS3-39b and provides an extra 74 bp of 5' pFS3 sequence. The second contains four of the inserts, pFS3-26a, pFS3-49a, pFS3-49b and pFS3-59a, all of which provide an extra 78 bp of sequence. The final group, containing the pFS3-26b insert, extends the pFS3 clone sequence by 81 bp and is the largest 5'-RACE product which

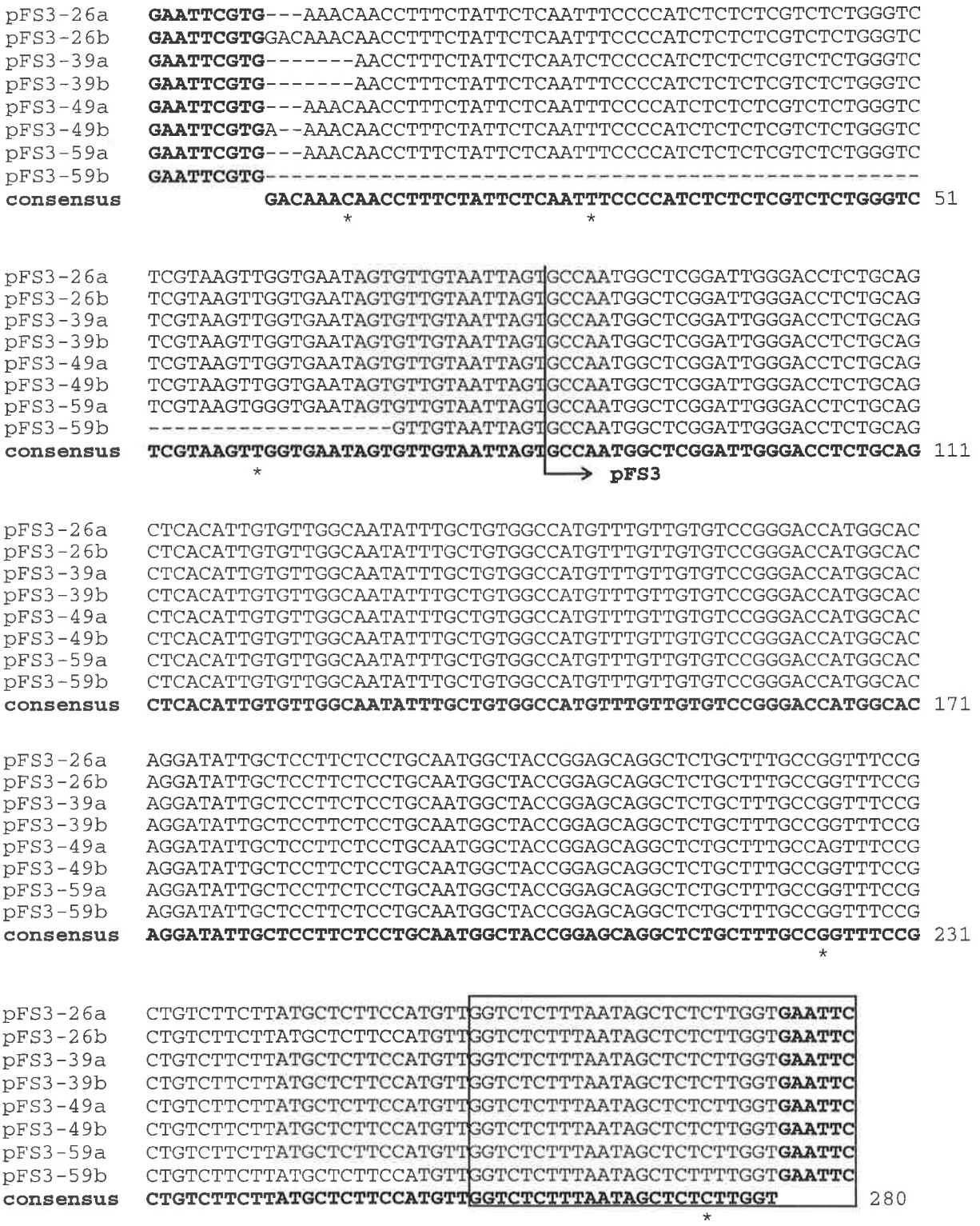


Figure 4.10: pFS3 5'-RACE clone sequences

Alignment of the eight insert sequences derived from pFS3 5'-RACE clones pFS3-26, pFS3-39, pFS3-49 and pFS3-59, together with a consensus sequence. The sequence of 5'-RACE PCR primer P3.2 is boxed and *Eco*RI restriction sites 5'-GAATTC-3' are in bold, as are the anchor sequences 5'-GTG-3'. The 5' end of the pFS3 sequence is indicated by an arrow, and nucleotide differences between the 5'-RACE inserts are indicated by asterisks below the consensus sequence. Numbers refer to the consensus sequence, which was constructed using the alignment of the eight sequences together with the 5' end of the longest clone, pFS3-26b.

was cloned in this experiment. The length variants may have arisen by premature termination of the reverse transcriptase, mRNA degradation, or by degradation of the cDNA product at any number of stages in the 5'-RACE procedure. Alternatively, the variant cDNAs may represent distinct pFS3 transcripts from the same gene or from two of more members of a pFS3 gene family.

4.5.2d The Full-length pFS3 cDNA

The sequence of the full-length cDNA, designated cFS3, was constructed using the consensus RACE clone sequence (Figure 4.10) combined with the pFS3 cDNA clone sequence (Figure 4.8) and is shown in figure 4.11. The sequence was 621 bp in length, including the poly(A) tail of 21 bp. The extra sequence information provided by the 5'-RACE clones does not contain any AUG initiation codons. The longest open reading frame identified in the three forward reading frames, termed ORF1, encodes a polypeptide of 66 amino acids with a calculated molecular weight of 6553 Da and a pI of 7.5 (Figure 4.11). A hydropathy profile (Kyte and Doolittle, 1982) of ORF1 revealed the generally hydrophobic nature of the protein and the conceptual amino acid sequence showed a predominance of alanine residues, constituting 21% of the protein. Neither of the two initiation codons identified in the cFS3 sequence conform to a consensus sequence which surrounds initiation codons (Kozak, 1984; Lütcke *et al.*, 1987). The longest protein possibly encoded by cFS3, ORF2, has a conceptual amino acid sequence which extends from the beginning of the cDNA sequence and terminates after the stop codon of ORF1 (Figure 4.11). ORF2 lacks an initiation codon in the cFS3 sequence, implying that the full-length pFS3 5'-RACE product has yet to be isolated. This is consistent with the poor correlation between an estimated pFS3 transcript size of 700 nt and a putative full-length cDNA of 621 bp. The 103 amino acid polypeptide encoded by ORF2 has a calculated molecular weight of 11,122 Da and a pI of 8.3, with no distinguishing features in its amino acid composition.

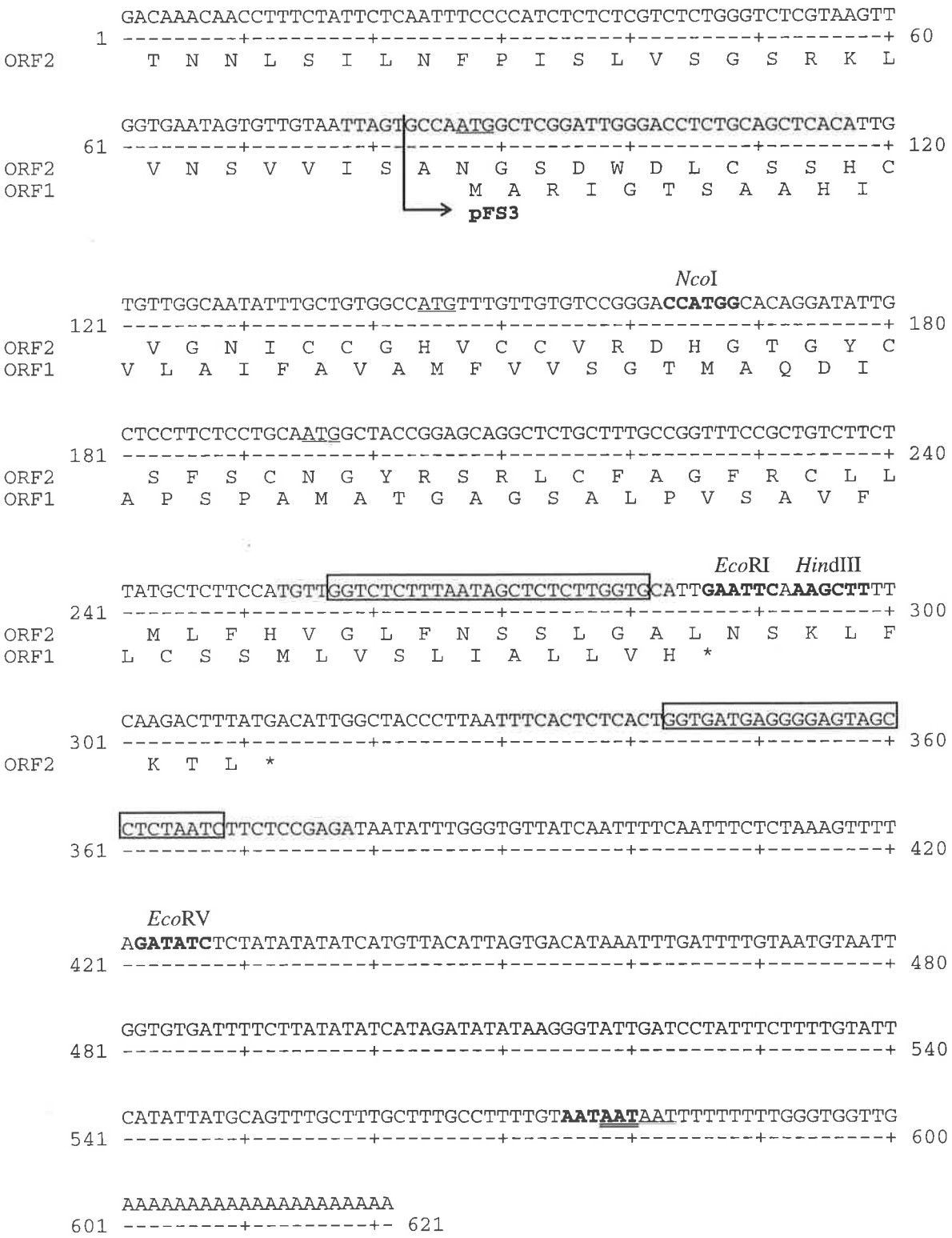
Neither of the conceptual translation sequences showed homology to sequences in the protein databases (PIR, SwissProt and GenPept, September 1994). Furthermore, the sequences lacked any amino acid patterns (motifs) which may have provided a suggestion as to their possible function(s) in the cotton fibre. Small regions of sequence derived from the three forward reading frames of cFS3 showed some similarity to the H^+-transporting ATP synthase from *Trypanosoma brucei* (Corell *et al.*, 1994) but the best of the alignments (ten identical residues and seven conserved residues in an overlap of 17 amino acid residues) has an associated probability of 11% that the similarity is due to chance (Altshul *et al.*, 1990; Gish and States, 1993). In addition, corresponding nucleotide similarity was not observed, with no substantial similarity found to cFS3 by any sequences in the nucleotide databases (GenBank and EMBL, September 1994). The 3' untranslated regions associated with ORF1 and ORF2 are 314 bp and 289 bp in length respectively, both of which seem unusually large.

There are several possible conclusions which may be drawn from the sequence analysis of fibre-specific clone pFS3. Firstly, the full-length cDNA clone may have been isolated and it encodes a novel protein product of 66 amino acids defined by ORF1. Whilst this seems short for a mature protein, genes encoding metallothionein proteins which range in size from 45 amino acids to 81 amino acids have been recently isolated from *Arabidopsis* (Zhou and Goldsbrough, 1995). Moreover, an 88 amino acid protein has been isolated from tomato and appears to represent a novel class of cell wall proteins (Domingo *et al.*, 1994). The protein is deposited specifically in cells containing lignified secondary cell walls such as tracheary elements, and may have a function in reinforcement of the walls of such cells. Secondly, it may be that the full-length cDNA has been obtained but there are errors in the sequence which result in disruption of the reading frame which contains ORF1 and formation of stop codon(s). Such errors may result in a truncated open reading frame or a conceptual translation which is derived from an incorrect reading frame and may have occurred as a result of misincorporation by reverse transcriptase. The possibility of sequencing errors also cannot be discounted. In

Figure 4.11: Nucleotide sequence of full-length cDNA clone cFS3

Sequence of the putative full-length fibre-specific clone cFS3, constructed from the RACE clone consensus sequence and that of the pFS3 cDNA. The 5' end of the pFS3 sequence is indicated by an arrow and the target sites for restriction enzymes *Nco*I, *Eco*RI, *Hin*dIII and *Eco*RV are shown in bold and are at nucleotide positions 163, 286, 293 and 422 respectively. The first putative polyadenylation signals is in bold and the second, which overlaps the first, is double underlined and the regions complementary to the oligonucleotide primers used for 5'-RACE are boxed.

Possible ATG initiation codons at the 5' end of the sequence are underlined and the predicted sequences of the longest open reading frames, ORF1 and ORF2, are shown in single letter code below the cDNA sequence. Stop codons are shown by asterisks.


GACAAACAACCTTTCTATTCTCAATTTCCCCATCTCTCTCGTCTCTGGGTCTCGTAAGTT
1
60
ORF2 T N N L S I L N F P I S L V S G S R K L
GGTGAATAGTGTTGTAATTAGTGCCAATGGCTCGGATTGGGACCTCTGCAGCTCACATTG
61
120
ORF2 V N S V V I S A N G S D W D L C S S H C
ORF1 M A R I G T S A A H I
pFS3
NcoI
TGTTGGCAATATTTGCTGTGGCCATGTTTGTTGTGTCCGGGACCATGGCACAGGATATTG
121
180
ORF2 V G N I C C G H V C C V R D H G T G Y C
ORF1 V L A I F A V A M F V V S G T M A Q D I
CTCCTTCTCCTGCAATGGCTACCGGAGCAGGCTCTGCTTTGCCGGTTTCCGCTGTCTTCT
181
240
ORF2 S F S C N G Y R S R L C F A G F R C L L
ORF1 A P S P A M A T G A G S A L P V S A V F
EcoRI HindIII
TATGCTCTTCCATGTTGGTCTCTTTAATAGCTCTCTTGGTGCATTGAATTCAAAGCTTTT
241
300
ORF2 M L F H V G L F N S S L G A L N S K L F
ORF1 L C S S M L V S L I A L L V H *
CAAGACTTTATGACATTGGCTACCCTTAATTTCACTCTCACTGGTGATGAGGGGAGTAGC
301
360
ORF2 K T L *
CTCTAATCTTCTCCGAGATAATATTTGGGTGTTATCAATTTTCAATTTCTCTAAAGTTTT
361
420
EcoRV
AGATATCTCTATATATATCATGTTACATTAGTGACATAAATTTGATTTTGTAATGTAATT
421
480
GGTGTGATTTTCTTATATATCATAGATATATAAGGGTATTGATCCTATTTCTTTTGTATT
481
540
CATATTATGCAGTTTGCTTTGCTTTGCCTTTTGTAATAATAATTTTTTTTTGGGTGGTTG
541
600
AAAAAAAAAAAAAAAAAAAA
601
621

either of these scenarios, the pFS3 transcript length, estimated from a Northern blot, is an overestimation and the most likely transcript size is nearer to the cFS3 sequence length (621 bp) than to 700 nt.

Another possibility is that cFS3 does not represent the full-length pFS3 mRNA. The full-length 621 bp cFS3 sequence is not consistent with the length of the pFS3 mRNA detected on RNA gel blots (Figure 4.4a) and this result suggests that the full-length pFS3 transcript has yet to be isolated. However, a second 5'-RACE experiment, in which the cDNA extension time was increased, did not increase the size of the 5'-RACE product (not shown). In this case, the 5' end of the cFS3 sequence is missing and the correct reading frame, encompassing ORF2, may encode a novel protein product.

Another explanation and the most plausible, is that the pFS3 cDNA clone is chimeric and consists of two or more different cDNAs which ligated together. However, despite presence of an *Eco*RI site at nucleotide 204 of the pFS3 sequence (Figure 4.8; Figure 4.11), there was no evidence of the cDNA linker sequence (5'-GAATTCGCGGCCGC-3') or of an adjacent *Not*I site to indicate ligation of two independent cDNAs. Internal poly(A) tails were also not apparent from the pFS3 sequence. It is possible that two inserts, which are rendered blunt-ended in the cDNA synthesis reaction, combined to form clone pFS3. In this case, the extreme 3' end of cFS3 may correspond to pFS3, but the 5' end, including the P3.1 and P3.2 5'-RACE primers, is an unrelated sequence which does not affect the fibre-specific hybridisation pattern of the pFS3 probe (Figure 4.4a). The conceptual amino acid sequence derived from such a clone, if any substantial ORFs were present, would be chimeric, possibly out of frame, and may not contain homology to known protein sequences.

Further experiments would involve isolation and sequence analysis of fibre-specific pFS3 homologues from the 13 DPA cotton fibre cDNA library. Comparison of the resulting sequences with that of pFS3 would identify region(s) of the latter clone which represent the fibre-specific transcript. A pair of 5'-RACE primers designed to anneal to this sequence would

result in the isolation of the full-length fibre-specific transcript.

4.5.2e Southern Analysis of Fibre-specific Clone pFS3

Southern analysis of pFS3 revealed a simple organisation of pFS3 genomic sequences, with a maximum of two copies of the pFS3 gene in the cotton tetraploid genome (Figure 4.12). Restriction of genomic DNA with enzymes which do not cut the pFS3 insert (*Bam*HI, *Bgl*II, *Xba*I, *Sac*I and *Xho*I) resulted in one or two hybridising fragments per sample of genomic DNA (Figure 4.12, lanes 2, 3, 6, 8 and 9). Enzymes which do cut within the pFS3 sequence (Figure 4.11) generally produced at least two genomic DNA fragments with homology to pFS3 (Figure 4.12, lanes 1, 4, 5 and 7).

Screening of a 1.6 x 10^5 clone Siokra 1-2 partial genomic library with the pFS3 insert did not identify any positive plaques. This is not unexpected since, based on a cotton tetraploid genome size of 3.5 x 10^6 kb (Geever, 1980), 8.5 x 10^6 clones would be required to give a 99% chance of identifying a particular single-copy sequence (Clarke and Carbon, 1976).

4.5.3 The pFS17 Clone Group

4.5.3a Initial Characterisation of pFS17 Homologues

Three fibre-specific cDNA clones were shown to contain sequences homologous to the representative and shortest clone, pFS17. In addition to pFS17, inserts from clones pFS12 and pFS14 hybridised to a pFS17-derived probe (Figure 4.3b). Clone pFS12 contained a very short insert with an estimated size of approximately 150 bp (Figure 4.3a) but the insert of pFS14, on the other hand, was the largest of the 24 fibre-specific cDNA clones and was estimated at approximately 1.5 kb in length (Figure 4.3a). Given an estimated transcript size for pFS17 of 1500 nt (Figure 4.4), one might expect that clone pFS14 contains the full-length pFS17 cDNA. The inserts of pFS12 and pFS17 were sequenced in their entirety, yielding 112 bp and 160 bp respectively (Figure 4.13a). Comparison of the two sequences revealed that the sequences were identical except that pFS12 was truncated at the 3' end and four of the five 3'-terminal

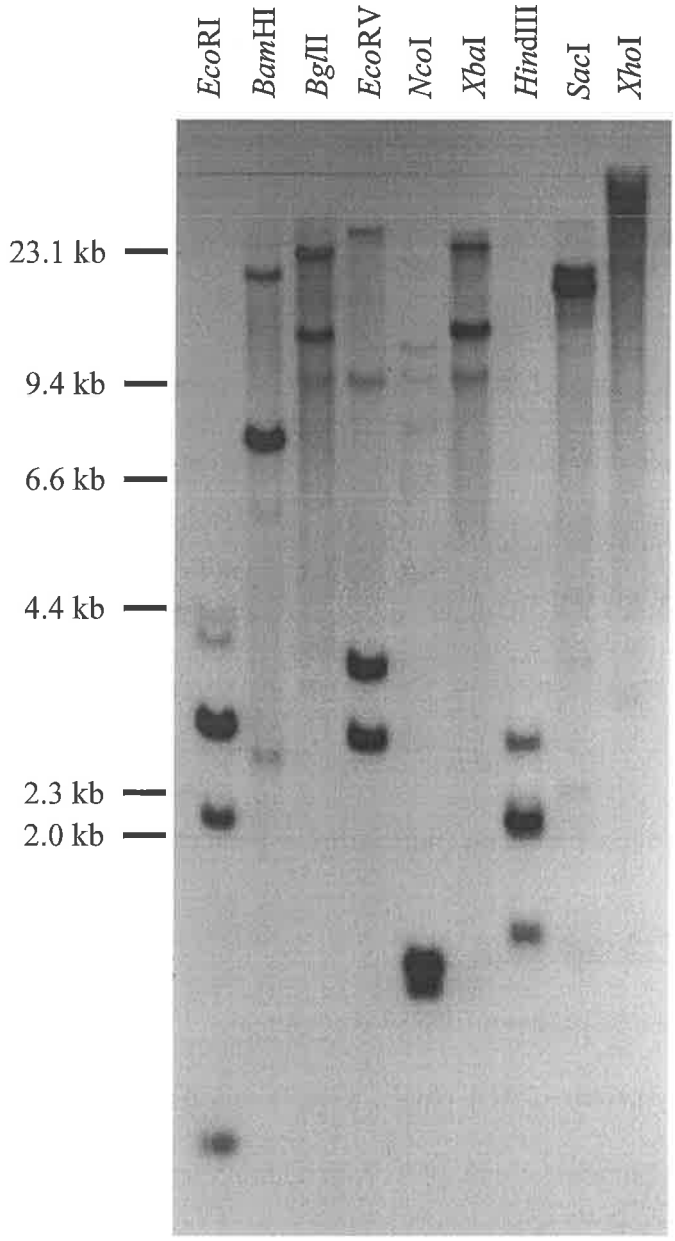


Figure 4.12: Southern analysis of fibre-specific cDNA clone pFS3

Genomic Southern blot probed with the pFS3 insert. Each lane was loaded with 10 µg of total genomic DNA from *G. hirsutum* L., cv. Siokra 1-2 cut with enzymes which recognise 6 bp target sites, as indicated. Numbers on the left represent the molecular weights of wild type Lambda DNA restricted with *Hin*dIII (kb).

nucleotides diverged in sequence from that of pFS17. The basis for this divergence is unknown, but the lack of poly(A) tail on clone pFS12 is probably due to premature termination of second-strand cDNA synthesis, as discussed in Section 3.3.

The insert of pFS14 was sequenced from both ends and as expected, the resultant sequences did not overlap. Sequence from the RSP of pBluescript®SK(-) yielded 386 bp of pFS14 insert sequence, of which 101 bp showed sequence similarity to the inserts of pFS12 and pFS17 (Figure 4.13b). The sequences of pFS17 and pFS14R were identical except that pFS14R appeared truncated at the 3' end of the cDNA, probably terminating at nucleotide 101 of the pFS17 sequence (Figure 4.13a, b). Following this region was a stretch of 23 T residues, which is not part of the pFS17 sequence and cannot constitute the poly(A) tail of the pFS14 transcript. It is clear from the sequence pFS14R that pFS14 is a chimeric clone which contains at least two independent cDNA inserts, and that the series of T residues corresponds to the poly(A) tail of a second insert, which is in opposite orientation to the pFS17-like insert. Sequence from the (-40) primer of pBluescript®SK(-) into the pFS14 insert yielded 429 bp of sequence (Figure 4.13c) which did not show any similarity to the pFS17 or pFS14R sequences.

Sequence information from the fibre-specific clone pFS17 was used in an attempt to determine potential function(s) of the encoded protein. However, the sequence of pFS17 identified no definitive homologies in the nucleotide databases (GenBank and EMBL, September 1994), probably due to incomplete sequence information. Similarly, the first 254 bp of the pFS14R sequence (Figure 4.13b) failed to identify any similar sequences in the databases. Use of the 132 bp pFS14 sequence upstream (3') of the pFS17 section to screen nucleotide databases (GenBank and EMBL, December 1994) was also inconclusive. Minor similarity was found to generally AT-rich sequences, such as those from mitochondrial genomes.

The 429 bp of sequence derived from the opposite end of pFS14 (pFS14F; Figure 4.13c)

Figure 4.13: Sequencing of the pFS17 group clones

Nucleotide sequences of the three pFS17 group cDNA clones. The 160 bp of pFS17 sequence includes a poly(A) tail of 24 adenine residues, and the pFS14R sequence included an internal poly(A) tail, but the sequence of clone pFS12 lacked a poly(A) tail.

a. Alignment of the sequences of cDNA clones pFS17 and pFS12, with single nucleotide differences indicated by asterisks below the alignment. A putative TAA stop codon at nucleotide position 24 in the pFS17 sequence is underlined and the regions complementary to the 5'-RACE cDNA primer (P17.1) and the 5'-RACE PCR primer (P17.2) overlap by two nucleotides and are boxed.

b. Partial sequence of cDNA clone pFS14 from the RSP of pBluescript®SK(-), designated pFS14R. The 101 bp of sequence which is identical to the sequence of pFS17 is in bold and underlined and extends from nucleotide 154 to nucleotide 254 of the pFS14R sequence.

c. Partial sequence of cDNA clone pFS14, from the (-40) primer of pBluescript®SK(-) and designated pFS14F. The conceptual amino acid sequence is shown in single letter code below the cDNA sequence and the putative site of signal sequence cleavage is indicated by an arrow. The three possible ATG initiation codons are underlined.

a.

```
          1 ---------+---------+---------+---------+---------+---------+ 60
pfs17       TTTTATTCTTGTATGGAATGTAGTAATGTTTTCTTTCATGTTTGAAGAGAGACCAGTTGC
pfs12     1 --------------GGAATGTAGTAATGTTTTCTTTCATGTTTGAAGAGAGACCAGTTGC 46
                                                  P17.2

         61 ---------+---------+---------+---------+---------+---------+ 120
pfs17       TCATGGCTGTACTCATTCGAATGCTTATTCTATATCTTCTTCTTTCTCAATGAGCAATAT
pfs12     57 TCATGGCTGTACTCATTCGAATGCTTATTCTATATCTTCTTCTTTCTCAATGAGCAATAT 106
                P17.1

        121 ---------+---------+---------+---------+ 160
pfs17       CCTGATGATTTCAAATAAAAAAAAAAAAAAAAAAAAAAA
pfs12   117 CGCGTA------------------------------- 112
             ** **
```

b.

```
    CTCCGGCACATGAGCCACCATACGGTCACTATCCAGGACATCCACCTTTGGGGAAGCCTC
  1 ---------+---------+---------+---------+---------+---------+ 60

    AATAGAATGCTCCCATCGACTAGCCATTCCCAACAATTGAGCTACTAGAAAAATTAGTAT
 61 ---------+---------+---------+---------+---------+---------+ 120

    TTATAATAAGAGGAACTGCCGGCTGATGTTCTGTTTTATTCTTGTATGGAATGTAGTAAT
121 ---------+---------+---------+---------+---------+---------+ 180

    GTTTTCTTTCATGTTTGAAGAGAGACCAGTTGCTCATGGCTGTACTCATTCGAATGCTTA
181 ---------+---------+---------+---------+---------+---------+ 240

    TTCTATATCTTCTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTATAAAAAAAGGTATTCAA
241 ---------+---------+---------+---------+---------+---------+ 300

    TAATGAAAACCATAAATATCATGGTCAATCATCAACAAACAAATTCAGACATTATGGATA
301 ---------+---------+---------+---------+---------+---------+ 360

    TATCCTCTAGAGTTCAAACTTAATAG
361 ---------+---------+------ 386
```

c.

```
    TTAGCTAGCTCTTACTCAAATGGCAACCAAAACGATGATGTTGCAAATATTTCCACTTTT
  1 ---------+---------+---------+---------+---------+---------+ 60
                        M  A  T  K  T  M  M  L  Q  I  F  P  L  F

    CTTCTTTTTGTTCAGTGTCTGCAACTCCATTTTCCTTGGTGCTAATGGAGATGACAATGG
 61 ---------+---------+---------+---------+---------+---------+ 120
      F  F  L  F  S  V  C  N  S  I  F  L  G  A  N  G  D  D  N  G

    TGGTTGGCAAACTGCCCATGCCACCTTCTACGGTGGTGCTGATGCTACCGGCACAATGGG
121 ---------+---------+---------+---------+---------+---------+ 180
      G  W  Q  T  A  H  A  T  F  Y  G  G  A  D  A  T  G  T  M  G

    GGGAGCTTGTGGTTATGGAAACCTGTACAGTCAAGGGTATGGAACGAGCACAGCAGCTTT
181 ---------+---------+---------+---------+---------+---------+ 240
      G  A  C  G  Y  G  N  L  Y  S  Q  G  Y  G  T  S  T  A  A  L

    GAGCACTGCACTTTTCAACAATGGCTTGAGCTGCGGTGCCTGCTACGAGCTCCGGTGCAA
241 ---------+---------+---------+---------+---------+---------+ 300
      S  T  A  L  F  N  N  G  L  S  C  G  A  C  Y  E  L  R  C  N

    CAATGATCCTCAATGGTGCATTAGTCGAACCATAACCGTGACAGCCACCAACTTTTGTCC
301 ---------+---------+---------+---------+---------+---------+ 360
      N  D  P  Q  W  C  I  S  R  T  I  T  V  T  A  T  N  F  C  P

    ACCTAACTATGCTTTATCTAGTGACAATGGCGGGTGGTGCAATCCCCACGAGAACACTTT
361 ---------+---------+---------+---------+---------+---------+ 420
      P  N  Y  A  L  S  S  D  N  G  G  W  C  N  P  H  E  N  T  L

    GATTGGCCG
421 --------- 429
      I  G
```

did not contain pFS17-like sequences, but showed significant homology to expansin genes from several plant species including *Arabidopsis*, cucumber, rice and pea (Newman *et al.*, 1994; Shcherban *et al.*, 1995). The best match was 70.2% nucleotide identity with the corresponding 379 bp region of an expansin gene recently identified in cucumber (Shcherban *et al.*, 1995). An incomplete ORF was identified in reading frame two of the pFS14F nucleotide sequence (Figure 4.13c), with three possible AUG initiation codons. The remaining two reading frames contained several stop codons. The position of the functional initiating codon is unknown but is assumed to be the first AUG, at nucleotide position 20 in the pFS14F sequence, since nucleotides in the vicinity of the start codon best match the consensus translation start site 5'-TAAACAATGGCT-3' (Joshi, 1987b) and the "first AUG" rule holds for the vast majority of plant genes (Kozak, 1984; Joshi, 1987b). The conceptual amino acid sequence of 135 residues showed significant similarity to the N-terminal sequences of expansin proteins from cucumber (Shcherban *et al.*, 1995), *Arabidopsis* (Shcherban *et al.*, 1995), rice (Shcherban *et al.*, 1995) and pea (Michael A J, unpublished) with amino acid identity of 75%, 65%, 72%, 73%, 72%, 70% and 69% respectively (Figure 4.14). The alignment shown in Figure 4.14 constitutes approximately half of the mature expansin proteins, since they attain a length of around 230 amino acid residues (Shcherban *et al.*, 1995). Amino acid similarity decreased markedly over the seven C-terminal amino acids of the pFS14F sequence (Figure 4.14) and may indicate presence of a sequencing error in the latter part of the sequence, or junction with another, unrelated cDNA.

Analysis of the predicted amino acid sequence of pFS14F revealed an N-terminal region with several of the features characteristic of eukaryote signal peptides (Nothwehr and Gordon, 1990). Namely, signal sequences consist of a highly hydrophobic central core of 7-16 amino acids, flanked by charged residues and with a short block of relatively polar residues which immediately precede the cleavage site. They may be 20-40 amino acids in length. The cleavage site typically has a bulky aromatic residue at the -2 position relative to the cleavage site, and

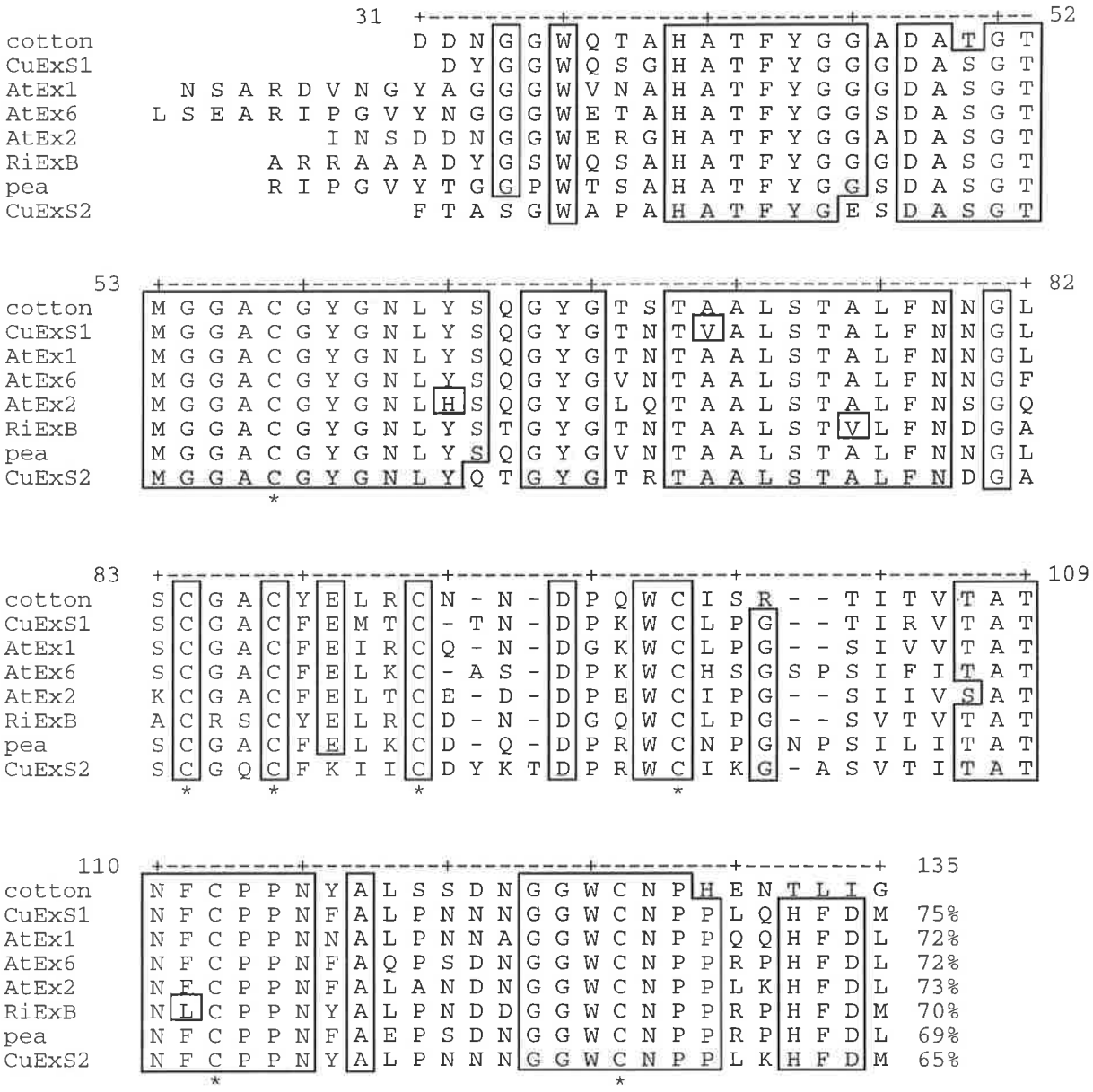


Figure 4.14: Analysis of a putative expansin from cotton

Amino acid sequence alignment of a putative cotton expansin with the corresponding regions of seven expansin-like proteins from cucumber (CuExS1, CuExS2), *Arabidopsis* (AtEx1, AtEx2, AtEx6), rice (RiExB) and pea. Mature proteins, after signal peptide cleavage, are shown and numbering refers to the cotton expansin preprotein sequence. Boxes indicate homologous amino acids in at least seven of the sequences and identity with the deduced amino acid sequence of pFS14F is indicated at the end of each sequence. Seven of the eight highly conserved cysteines present in the N-terminal half of expansin proteins are contained within the pFS14R conceptual translation and are denoted by asterisks.

small, neutral residues at amino acid positions -1 and -3 in the signal sequence (von Heijne, 1986). Proline residues are not found between residues -3 and +1 of the cleavage site. The predicted signal sequence cleavage site was obtained according to von Heijne (1983) and occurs between amino acids 30 and 31 of the protein encoded by pFS14R (Figure 4.13c). In plants, the presence of the signal sequence, together with absence of an endoplasmic reticulum signal KDEL or HDEL (Denecke *et al.*, 1992) is sufficient to direct the protein to the extracellular matrix *via* the endoplasmic reticulum-Golgi secretory pathway. Secretion of the cotton fibre expansin is consistent with previous findings that expansins are associated with plant cell walls (McQueen-Mason *et al.*, 1992) and is in agreement with a proposed role in cell growth.

Expansins are a unique class of cell wall proteins first isolated in 1992 from cucumber hypocotyls (McQueen-Mason *et al.*, 1992). Extensive and convincing studies by Cosgrove and coworkers (reviewed in McQueen-Mason, 1995) have elucidated the role of expansins as "wall-loosening" enzymes. Their putative mode of action involves a novel non-hydrolytic mechanism which disrupts the hydrogen-bonding between the xyloglucan backbone of the cell wall and the cellulose microfibrils (Cosgrove, 1993; Fry, 1994; McQueen-Mason and Cosgrove, 1995). The resultant relaxation of the cell wall polymeric network facilitates the turgor-driven expansion of plant cells (Section 1.3.3b). Phylogenetic analysis of expansins suggest that they constitute an ancient multigene family that arose before the evolutionary divergence of monocotyledons and dicotyledons (Shcherban *et al.*, 1995). Although the exact mechanism of action of expansins is unknown, functional regions have been inferred from conserved sites in the protein sequence and the spacing of eight conserved cysteines in the N-terminal half of the protein (Figure 4.14) suggests a particular folding configuration for expansin proteins (Wright *et al.*, 1991; Shcherban *et al.*, 1995). Expansin-like proteins are expected to be important in the development of such cells as cotton fibres, which undergo a rapid rate of cell elongation and, despite the fact that the cotton expansin cDNA which forms

part of the pFS14 clone may not be differentially expressed, the nature of the sequence warrants further investigation.

The region upstream (5') of the pFS17-homologous section of pFS14R may be further pFS17 sequence, or it may be sequence derived from another cDNA insert which blunt-end ligated to the pFS17 insert during the cDNA synthesis procedure. The remaining insert(s) do not affect the fibre-specificity of pFS14 as a whole, since the clone was hybridised only by fibre cDNA in the differential screen (not shown). The pFS17-homologous region is clearly sufficient to confer the fibre-specificity of pFS14, since only 112 bp of pFS12 was sufficient to confer fibre-specificity on the latter clone. Additional inserts may also encode fibre-specific cDNAs but this is unlikely given the rarity of fibre-specific cDNAs, and a more plausible explanation is that they encode low-level transcripts in all cotton tissues and therefore did not affect the isolation of pFS14 by differential screening.

4.5.3b Isolation of the Full-length pFS17 cDNA by 5'-RACE

Despite the possibility that clone pFS14 contained an extended pFS17 homologue of 254 bp, the only region which can be confirmed as encoding a fibre-specific cDNA is that 101 bp which was identical in sequence to pFS17 and pFS12 (Figure 4.13b). In any case, pFS14 cannot contain the full-length pFS17 cDNA, since the corresponding transcript size was estimated at 1500 nt (Figure 4.4).

5'-RACE was carried out in order to clone the remainder of the pFS17 transcript, using oligonucleotide primers designed to the pFS17 sequence (Figure 4.13a; Table 2.3b). 5'-RACE PCR generated a single amplified DNA fragment of approximately 1 kb in size, superimposed upon a smear of PCR products (Figure 4.15a, lane 3). A 5'-RACE PCR product of 1.1 kb does not reconcile with an estimated transcript length of 1500 nt, given that primer P17.2 is located 99 bp from the presumed 3' terminus of the pFS17 mRNA (Figure 4.13a). It may be that the transcript size was overestimated from the Northern blot (Figure 4.4b) in a similar way to that

Figure 4.15: Cloning of the 5' end of cDNA clone pFS17

a. Agarose gel showing products resulting from PCR amplification of 5'-RACE products of pFS17. Lanes 2-5 are each loaded with a 10 µl sample of PCR product, from a reaction volume of 50 µl and the lanes are:

1. Molecular weight markers (pUC19 DNA restricted with *Hpa*II) ,with fragment sizes 501 bp, 489 bp, 404 bp 331 bp, 242 bp, 190 bp, 147 bp and 111 bp;
2. TFR control cDNA after both purification and anchor ligation, utilising one gene-specific primer together with the anchor primer, which amplify a product of 480;
3. 5'-RACE PCR of pFS17 cDNA, utilising gene-specific PCR primer P17.2 together with the anchor primer;
4. Negative control reaction, using the same primers as those used in the experiment, but without the addition of a cDNA template;
5. A 1 µg sample of uncut pFS17 plasmid DNA, included as a positive control for the subsequent Southern analysis.

b. Southern blot of the gel in part a., hybridised with the pFS17 cDNA insert. Lanes are numbered as in part a.

c. Agarose gel showing products resulting from modification of 5'-RACE PCR reaction conditions. Lanes 2-10 are each loaded with a 10 µl sample of PCR product, from a reaction volume of 50 µl and the lanes are:

1. A combination of molecular weight markers (Lambda DNA restricted with *Hin*dIII), with fragment sizes 23.1 kb, 9.4 kb, 6.6 kb, 4.4 kb, 2.3 kb and 2.0 kb, and pUC19 DNA restricted with *Hpa*II ,with fragment sizes 501 bp, 489 bp, 404 bp 331 bp, 242 bp, 190 bp, 147 bp and 111 bp;
2. PCR reaction using normal reaction conditions, as for part a.;
3. PCR reaction using normal reaction conditions, together with a decreased amount of template cDNA;
4. PCR reaction with an increased primer annealing temperature of 65°C;
5. PCR reaction with an increased primer annealing temperature of 65°C, together with a decreased amount of template cDNA;

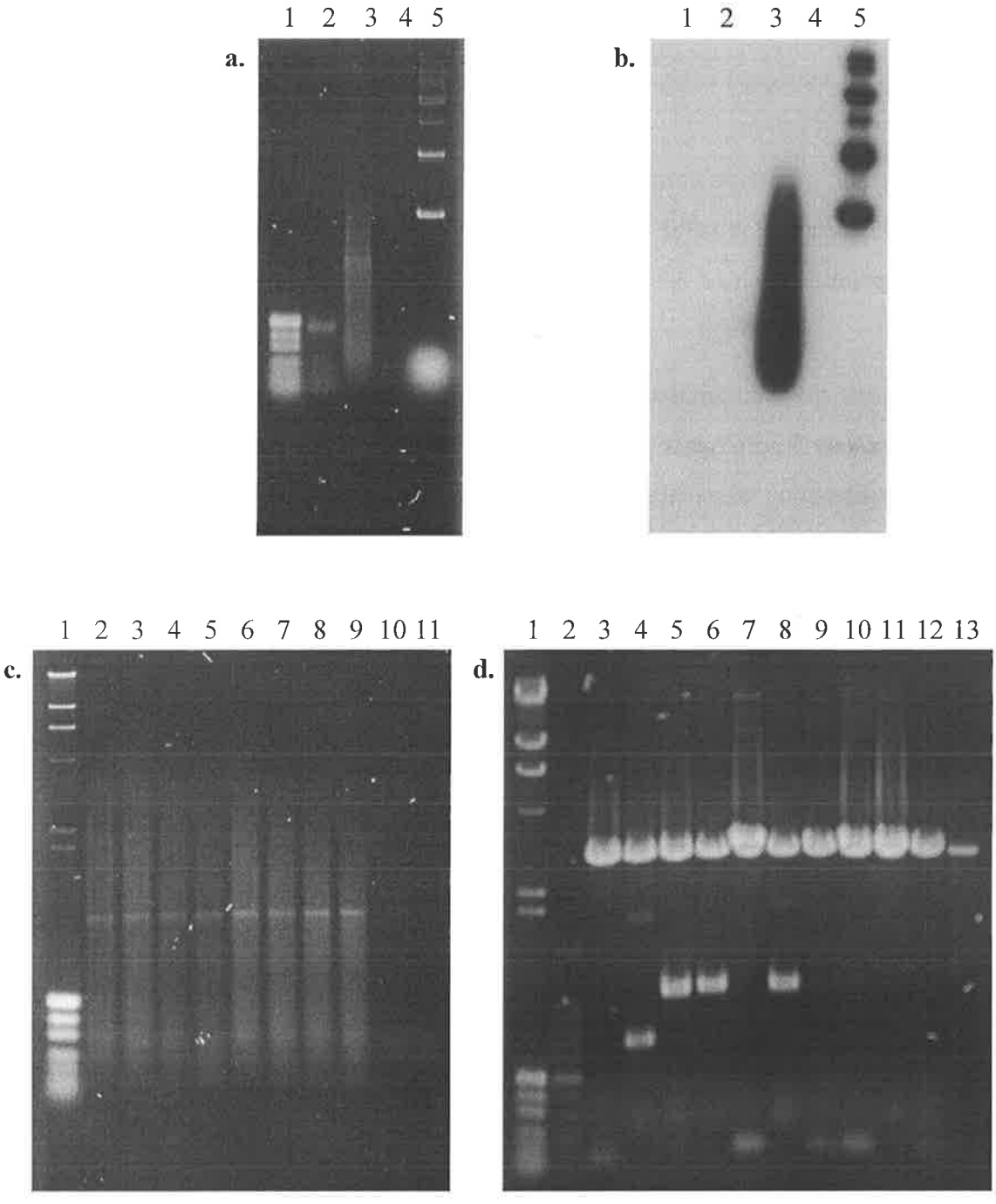
6. PCR reaction with an increased extension time of 3 minutes;
7. PCR reaction with an increased extension time of 3 minutes, together with a decreased amount of template cDNA;
8. PCR reaction with an increased primer annealing temperature of 65°C, together with an increased extension time of 3 minutes;
9. PCR reaction with an increased primer annealing temperature of 65°C, in combination with an increased extension time of 3 minutes and a decreased amount of template cDNA;
10. and 11. Negative control reactions, using the same primers as those used in the experiment and reaction conditions as for lane 6, but without the addition of a cDNA template.

d. Agarose gel electrophoresis of putative pFS17 5'-RACE clones. Each lane is loaded with approximately 2 µg of plasmid DNA restricted with *Eco*RI and the lanes are:

1. A combination of molecular weight markers (Lambda DNA restricted with *Hin*dIII), with fragment sizes 23.1 kb, 9.4 kb, 6.6 kb, 4.4 kb, 2.3 kb and 2.0 kb, and pUC19 DNA restricted with *Hpa*II ,with fragment sizes 501 bp, 489 bp, 404 bp 331 bp, 242 bp, 190 bp, 147 bp and 111 bp;
2. Molecular weight markers (100 bp DNA ladder), consisting of fragments that range in size from 100 bp to 1000 bp in 100 bp increments, plus an additional fragment of 1500 bp;

3.- 12. Putative pFS17 5'-RACE clones A-J;

13. A 1 µg sample of pBluescript®SK(-), restricted with *Eco*RI.


a.
1 2 3 4 5
b.
1 2 3 4 5
c.
1 2 3 4 5 6 7 8 9 10 11
d.
1 2 3 4 5 6 7 8 9 10 11 12 13

proposed for pFS3. Absence of products in the control reaction which lacked DNA (Figure 4.15a, lane 4) showed that the experimental reaction was specific and contained no contaminant DNA. PCR of positive control cDNA resulted in a single amplification product of the expected size, approximately 500 bp (Figure 4.15a, lane 2). Southern analysis verified the identity of the pFS17 5'-RACE PCR product (Figure 4.15b). The pFS17 probe hybridised strongly to the pFS17 5'-RACE product and the background smear in addition to uncut pFS17 plasmid. Hybridisation of probe pFS17 to the heterogeneous population of PCR products showed that the latter was not the result of non-specific priming of cDNA synthesis from diverse mRNAs.

The basis for the smearing and for the generation of PCR products longer than expected is unknown. The annealing temperature for the cDNA primer is low (cDNA synthesis occurs at 52^{o}C, whilst the calculated annealing temperature for P17.1 is 67^{o}C) so priming of cDNA synthesis may occur from several regions in the pFS17 transcript. Longer products may arise from ligation of several single-stranded cDNA products during the anchor ligation reaction. Absence of such a smear from 5'-RACE reactions of pFS3 (Section 4.5.2b) and pFS6 (Section 5.4) clones indicates that this is a phenomenon which is peculiar to the pFS17 cDNA sequence and the P17.1 and P17.2 5'-RACE primers. PCR parameters were altered in an attempt to optimise generation of the 1.1 kb product and minimise background products. The extension step of the PCR was increased from 2 minutes to 3 minutes in an effort to reduce the amount of any short PCR products, and the annealing temperature was increased from 60^{o}C to 65^{o}C (that calculated for primer P17.2) to minimise generation of non-specific products. The modifications, when used in combination with each other and with normal reaction conditions, did not appreciably improve the yield of the 1 kb product in comparison with the background smear (Figure 4.15c). Reduction in the amount of anchored cDNA template also did not improve the result, but a further reduction in the amount of ligation mix may have lessened the

smear somewhat (5'-AmpliFINDER™ RACE manual, CLONTECH). It may be that PCR conditions require further modification or that the smear products observed are generated in the cDNA reaction or prior to the PCR reaction.

The band of interest was excised from the agarose gel and the limited amount of insert was cloned into *Eco*RI-cut λZAP®II (Stratagene), yielding ten positive plaques of which three, pFS17-C, pFS17-D and pFS17-F contained inserts of the expected size (Figure 4.15d). The majority of clones contained small (approximately 100 bp) inserts and pFS17-B contained an insert of intermediate size, around 700 bp (Figure 4.15d, lane 4). All inserts hybridised strongly to a pFS17 probe in Southern blot analysis (not shown).

4.5.3c Sequence Analysis of Three pFS17 5'-RACE Clones

In a group of 18 restriction enzymes, only two, *Cla*I and *Kpn*I, contained recognition sites in 5'-RACE clone pFS17-C, and sites were located very close to the insert/vector junction (not shown). Subcloning was therefore difficult and the three pFS17 5'-RACE clones were sequenced using pBluescript®SK(-) primers, followed by oligonucleotide primers designed from the existing sequence. Primers P17R2, P17R3 and P17F2 (Table 2.3b) were designed to sequence derived from a single DNA strand and although efficient sequencing and/or PCR primers, subsequent sequencing revealed inaccuracies in primer design in the latter two cases. Clones pFS17-C, pFS17-D and pFS17-F yielded 1134 bp, 1142 bp and 1133 bp of sequence respectively, and an alignment of the three sequences, together with a consensus sequence, is shown in Figure 4.16. Initial sequencing of pFS17-C revealed a repeat structure which confounded primer design and further sequence analysis. Sequencing difficulties were resolved by PCR, in which the region of interest was subcloned by generation of a PCR product using an internal sequencing primer (Table 2.3b) together with the appropriate flanking primer, T3 or T7 (Table 2.3a).

As expected, the sequences of all three clones contained a 60 bp overlap with the

sequence of pFS17, to which they were identical. In addition, the consensus sequence (Figure 4.16) was identical to the corresponding region of pFS14 (Figure 4.13b), implying that pFS14R does indeed contain 254 bp of extended pFS17 sequence. Differences between the sequences, indicated by asterisks (Figure 4.16), can be attributed to *Taq* DNA polymerase errors during 5'-RACE PCR since in the nine cases, one clone mismatched the other two by a single nucleotide substitution.

The clones showed minor sequence and length heterogeneity at their 5' termini (Figure 4.16). All three clones contained *Eco*RI sites and 5'-GTG-3' anchor sequences as expected, and pFS17-C contained an additional anchor sequence and *Eco*RI site (Figure 4.16), presumably due to ligation of an anchor doublet with the cDNA during the 5'-RACE process. Clones pFS17-C and pFS17-F were identical at their 5' termini except for an additional G residue in pFS17-C. This probably arose by reverse transcription of the 5' cap-G structure into a terminal C residue, which gives rise to an uncoded G residue in the complementary strand sequence (Hirzmann *et al.*, 1993). The sequence of pFS17-D deviates from the other two clones by an omitted A residue and presence of an additional nine bp at the 5' end (Figure 4.16), the origin of which is unknown. It may be that pFS17-D encodes a full-length pFS17 transcript, whereas clones pFS17-C and pFS17-F are truncated, apparently at the same position. However, the extra sequence in clone D is considered to be a cloning artifact. A *Kpn*I site was detected near the 5' end of each cDNA, as expected from restriction mapping of clone pFS17-C, and a *Cla*I site was absent in the sequences of pFS17-D and pFS17-F but present in the extra anchor sequence which formed part of pFS17-C (Figure 4.16).

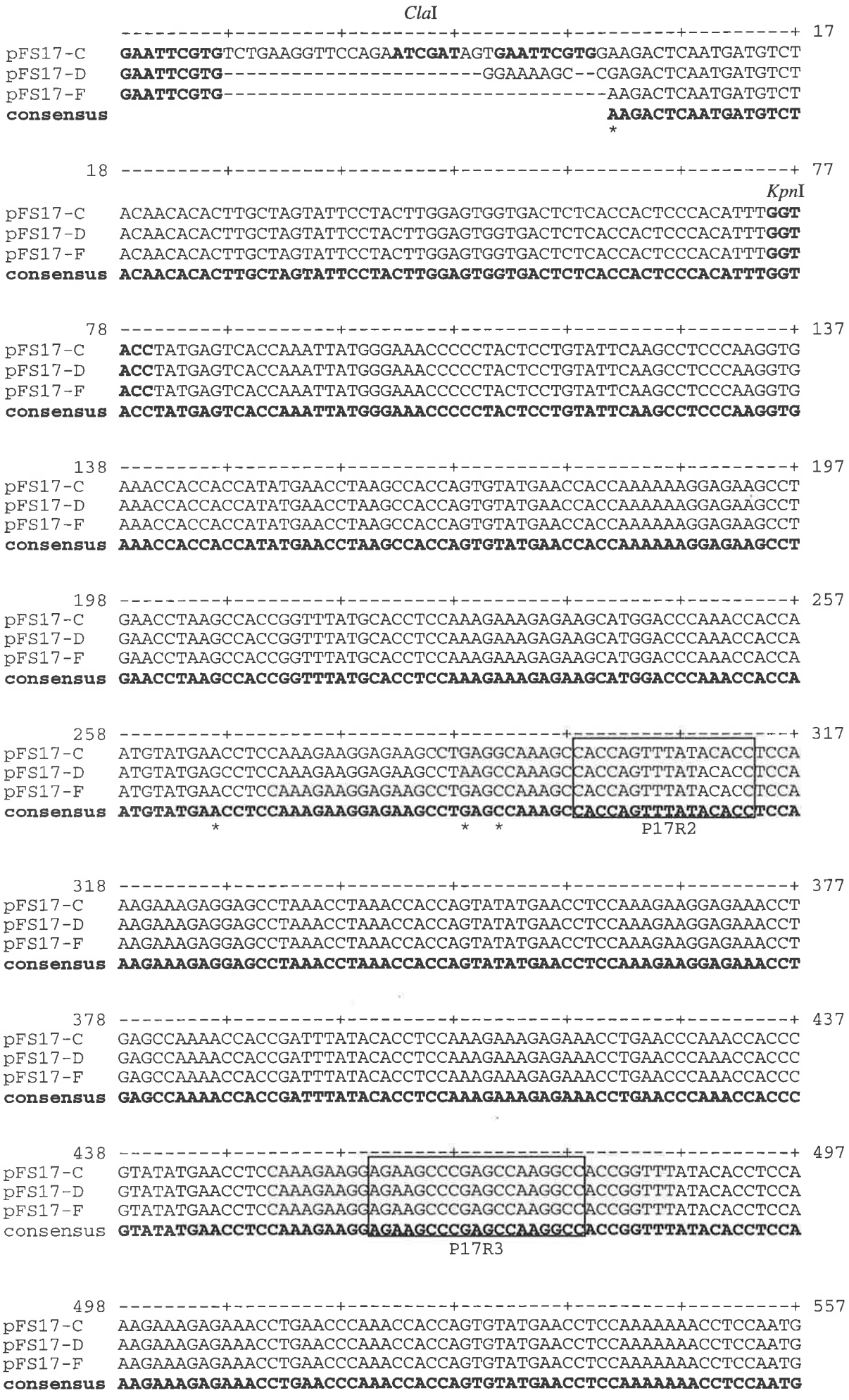
4.5.3d The Full-length pFS17 cDNA

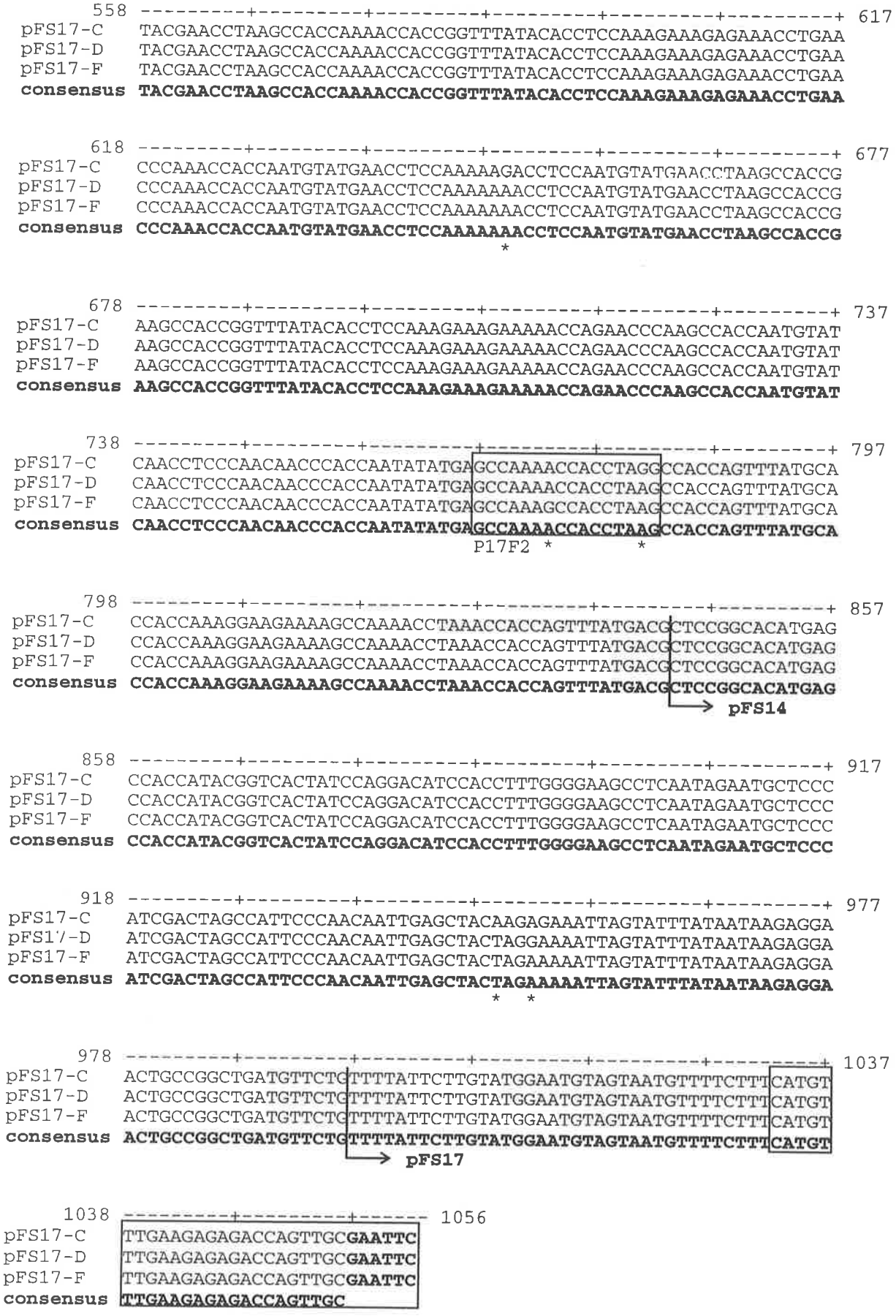
A 1132 bp sequence for the full-length pFS17 transcript, designated cFS17, was constructed from the 5'-RACE consensus sequence and the pFS17 sequence, and is shown in Figure 4.17. The repetitive nature of the sequence was clearly apparent and this repetition

Figure 4.16: Sequences of pFS17 5'-RACE clones

Alignment of the three insert sequences derived from pFS17 5'-RACE clones pFS317-C, pFS17-D and pFS17-F, together with a consensus sequence. The sequences corresponding to 5'-RACE PCR primer P17.2 and P17Γ2 are boxed, together with sequences complementary to the sequencing primers P17R2 and P17R3. The 5' ends of the pFS17 and pFS14 sequences are indicated by arrows. Differences between the sequences are indicated by asterisks below the consensus and numbering refers to the consensus sequence.

The *Kpn*I site 5'-GGTACC-3' (present in all clones and at nucleotide 75 in the consensus sequence) and the *Cla*I 5'-ATCGAT-3' site (present only in pFS17-C) are in bold. *Eco*RI restriction sites 5'-GAATTC-3' are also in bold, as are the anchor sequences 5'-GTG-3'.


ClaI
KpnI
P17R2
P17R3


558 617
pFS17-C TACGAACCTAAGCCACCAAAACCACCGGTTTATACACCTCCAAAGAAAGAGAAACCTGAA
pFS17-D TACGAACCTAAGCCACCAAAACCACCGGTTTATACACCTCCAAAGAAAGAGAAACCTGAA
pFS17-F TACGAACCTAAGCCACCAAAACCACCGGTTTATACACCTCCAAAGAAAGAGAAACCTGAA
consensus TACGAACCTAAGCCACCAAAACCACCGGTTTATACACCTCCAAAGAAAGAGAAACCTGAA
618 677
pFS17-C CCCAAACCACCAATGTATGAACCTCCAAAAAGACCTCCAATGTATGAACCTAAGCCACCG
pFS17-D CCCAAACCACCAATGTATGAACCTCCAAAAAAACCTCCAATGTATGAACCTAAGCCACCG
pFS17-F CCCAAACCACCAATGTATGAACCTCCAAAAAAACCTCCAATGTATGAACCTAAGCCACCG
consensus CCCAAACCACCAATGTATGAACCTCCAAAAAAACCTCCAATGTATGAACCTAAGCCACCG
678 737
pFS17-C AAGCCACCGGTTTATACACCTCCAAAGAAAGAAAAACCAGAACCCAAGCCACCAATGTAT
pFS17-D AAGCCACCGGTTTATACACCTCCAAAGAAAGAAAAACCAGAACCCAAGCCACCAATGTAT
pFS17-F AAGCCACCGGTTTATACACCTCCAAAGAAAGAAAAACCAGAACCCAAGCCACCAATGTAT
consensus AAGCCACCGGTTTATACACCTCCAAAGAAAGAAAAACCAGAACCCAAGCCACCAATGTAT
738 797
pFS17-C CAACCTCCCAACAACCCACCAATATATGAGCCAAAACCACCTAGGCCACCAGTTTATGCA
pFS17-D CAACCTCCCAACAACCCACCAATATATGAGCCAAAACCACCTAAGCCACCAGTTTATGCA
pFS17-F CAACCTCCCAACAACCCACCAATATATGAGCCAAAGCCACCTAAGCCACCAGTTTATGCA
consensus CAACCTCCCAACAACCCACCAATATATGAGCCAAAACCACCTAAGCCACCAGTTTATGCA
P17F2
798 857
pFS17-C CCACCAAAGGAAGAAAAGCCAAAACCTAAACCACCAGTTTATGACGCTCCGGCACATGAG
pFS17-D CCACCAAAGGAAGAAAAGCCAAAACCTAAACCACCAGTTTATGACGCTCCGGCACATGAG
pFS17-F CCACCAAAGGAAGAAAAGCCAAAACCTAAACCACCAGTTTATGACGCTCCGGCACATGAG
consensus CCACCAAAGGAAGAAAAGCCAAAACCTAAACCACCAGTTTATGACGCTCCGGCACATGAG
pFS14
858 917
pFS17-C CCACCATACGGTCACTATCCAGGACATCCACCTTTGGGGAAGCCTCAATAGAATGCTCCC
pFS17-D CCACCATACGGTCACTATCCAGGACATCCACCTTTGGGGAAGCCTCAATAGAATGCTCCC
pFS17-F CCACCATACGGTCACTATCCAGGACATCCACCTTTGGGGAAGCCTCAATAGAATGCTCCC
consensus CCACCATACGGTCACTATCCAGGACATCCACCTTTGGGGAAGCCTCAATAGAATGCTCCC
918 977
pFS17-C ATCGACTAGCCATTCCCAACAATTGAGCTACAAGAGAAATTAGTATTTATAATAAGAGGA
pFS17-D ATCGACTAGCCATTCCCAACAATTGAGCTACTAGGAAAATTAGTATTTATAATAAGAGGA
pFS17-F ATCGACTAGCCATTCCCAACAATTGAGCTACTAGAAAAATTAGTATTTATAATAAGAGGA
consensus ATCGACTAGCCATTCCCAACAATTGAGCTACTAGAAAAATTAGTATTTATAATAAGAGGA
978 1037
pFS17-C ACTGCCGGCTGATGTTCTGTTTTATTCTTGTATGGAATGTAGTAATGTTTTCTTTCATGT
pFS17-D ACTGCCGGCTGATGTTCTGTTTTATTCTTGTATGGAATGTAGTAATGTTTTCTTTCATGT
pFS17-F ACTGCCGGCTGATGTTCTGTTTTATTCTTGTATGGAATGTAGTAATGTTTTCTTTCATGT
consensus ACTGCCGGCTGATGTTCTGTTTTATTCTTGTATGGAATGTAGTAATGTTTTCTTTCATGT
pFS17
1038 1056
pFS17-C TTGAAGAGAGACCAGTTGCGAATTC
pFS17-D TTGAAGAGAGACCAGTTGCGAATTC
pFS17-F TTGAAGAGAGACCAGTTGCGAATTC
consensus TTGAAGAGAGACCAGTTGC

explains the lack of restriction enzyme sites within the 5'-RACE clones. Analysis of the pFS17 cDNA sequence between the putative stop codon and the polyadenylation site failed to detect a consensus polyadenylation signal, 5'-AATAAA-3'. However, putative polyadenylation sequences have been identified in several plant genes which diverge considerably from the animal consensus AATAAA and departure from the canonical sequence can be tolerated (Heidecker and Messing, 1986; Joshi, 1987a; Rothnie *et al.*, 1994). Several AATAAA-like motifs can be identified in the cFS17 3' untranslated region, one or more of which may be functional. The sequence 5'-AATAAG-3' is present 165 bp upstream of the polyadenylation site and the sequences 5'-AATGAG-3' and 5'-AATATC-3' are present 28 bp and 21 bp upstream of the cleavage site respectively (Figure 4.17).

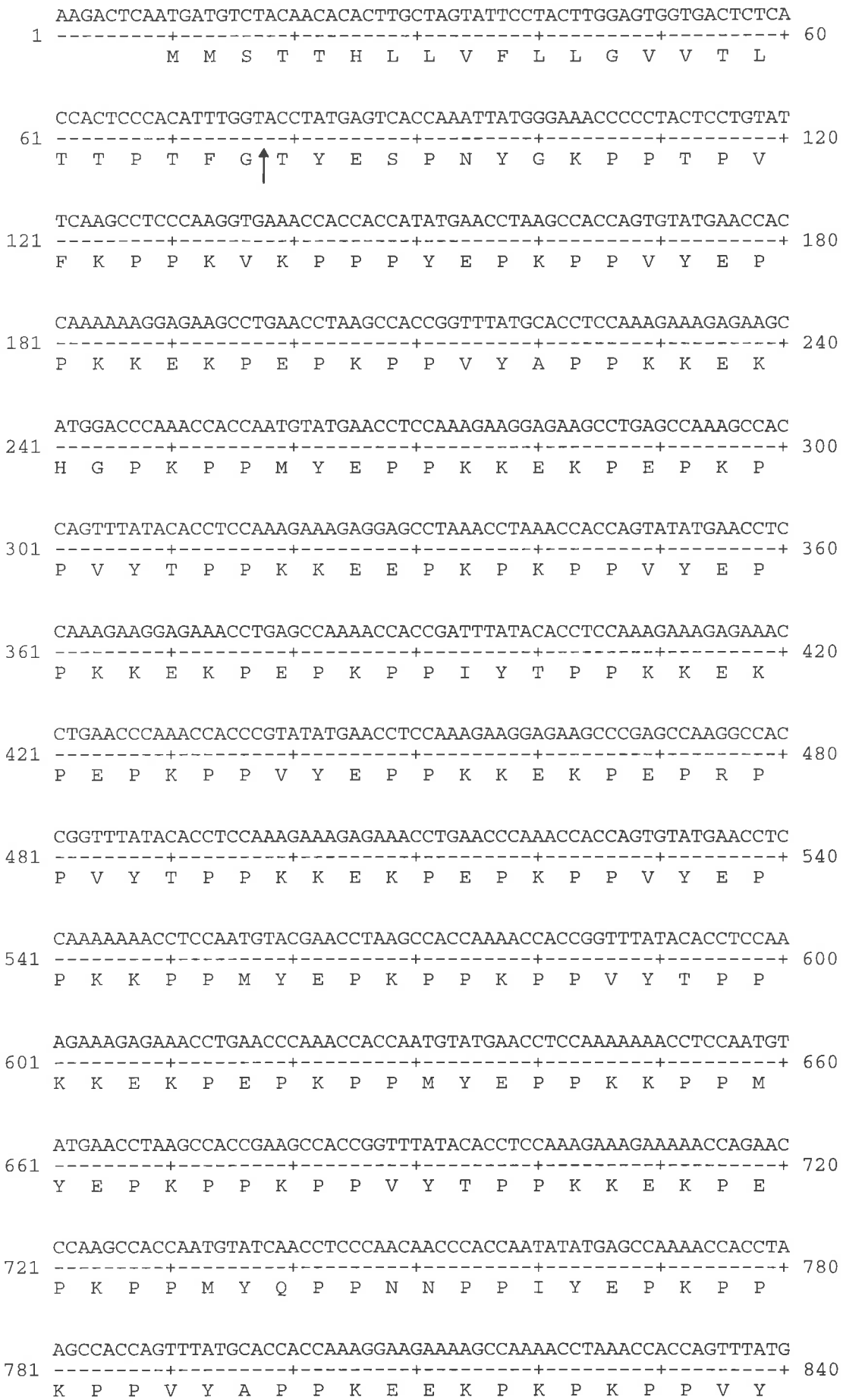
Several open reading frames were identified in the cFS17 sequence, the largest of which, designated preFSPRP, encodes a putative polypeptide of 299 amino acids (Figure 4.17) with a calculated molecular weight of 34028 Da and a pI of 10.5. The two possible start codons occur at nucleotide positions 9 and 12 in the cFS17 sequence and an appropriate stop codon, TAG, was present at nucleotide 906. The second of the two putative initiation codons in cFS17 is in optimum context. That is, an A residue is present at the -3 position relative to the initiation codon, and it best matches the consensus translation start site 5'-TAAACAATGGCT-3' (Joshi, 1987b). The first methionine codon is in suboptimum context, however the "first AUG" rule has been found to hold for approximately 92% of plant genes, as determined by Kozak (1984) and confirmed by Joshi (1987b) and the position of the functional initiating codon is assumed to occur at nucleotide position 9 in the cFS17 sequence.

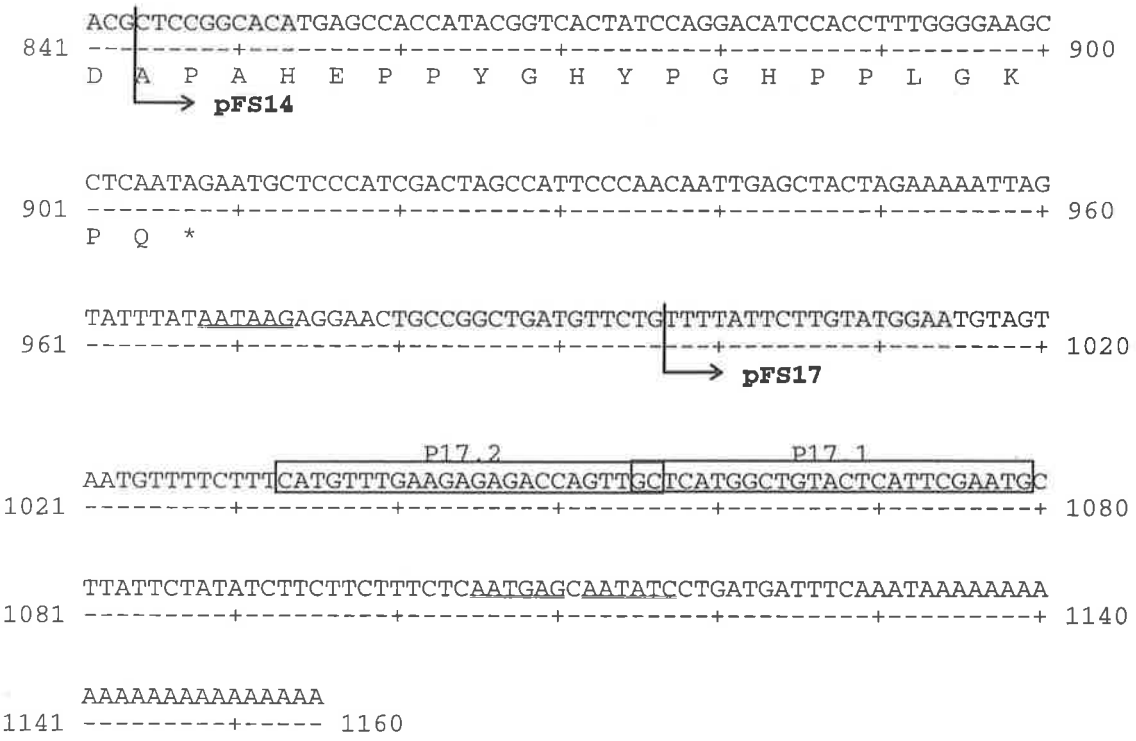
A comparison of the full-length pFS17 cDNA sequence with sequences in the databases (GenBank and EMBL, October 1994) revealed significant similarity at both the nucleotide and amino acid levels to plant proline-rich proteins (PRPs) including those isolated from *Medicago trunculata* (Wilson and Cooper, 1994; Wilson *et al.*, 1994), carrot (Ebener *et al.*, 1993), alfalfa (Löbler and Hirsch, 1993), tobacco (Keller and Lamb, 1989) and several soybean PRPs (Datta

Figure 4.17: Nucleotide sequence of full-length cDNA clone cFS17

Nucleotide sequence and conceptual translation of the putative full-length fibre-specific cDNA clone cFS17 (cFS17), constructed from the 5'-RACE clone consensus sequence and that of pFS17. The 5' ends of the pFS17 and pFS14 sequences are indicated by arrows and the regions complementary to the 5'-RACE cDNA primer (P17.1) and the 5'-RACE PCR primer (P17.2) overlap by two nucleotides and are boxed. Putative polyadenylation signals near the 3' terminus of the cDNA are double underlined.

The predicted sequence of the longest open reading frame, designated preFSPRP, is shown in single letter code below the cDNA sequence and the stop codon TAG is designated by an asterisk. The putative signal peptide cleavage site is marked with an arrow.


AAGACTCAATGATGTCTACAACACACTTGCTAGTATTCCTACTTGGAGTGGTGACTCTCA
1 60
M M S T T H L L V F L L G V V T L
CCACTCCCACATTTGGTACCTATGAGTCACCAAATTATGGGAAACCCCCTACTCCTGTAT
61 120
T T P T F G T Y E S P N Y G K P P T P V
TCAAGCCTCCCAAGGTGAAACCACCACCATATGAACCTAAGCCACCAGTGTATGAACCAC
121 180
F K P P K V K P P P Y E P K P P V Y E P
CAAAAAAGGAGAAGCCTGAACCTAAGCCACCGGTTTATGCACCTCCAAAGAAAGAGAAGC
181 240
P K K E K P E P K P P V Y A P P K K E K
ATGGACCCAAACCACCAATGTATGAACCTCCAAAGAAGGAGAAGCCTGAGCCAAAGCCAC
241 300
H G P K P P M Y E P P K K E K P E P K P
CAGTTTATACACCTCCAAAGAAAGAGGAGCCTAAACCTAAACCACCAGTATATGAACCTC
301 360
P V Y T P P K K E E P K P K P P V Y E P
CAAAGAAGGAGAAACCTGAGCCAAAACCACCGATTTATACACCTCCAAAGAAAGAGAAAC
361 420
P K K E K P E P K P P I Y T P P K K E K
CTGAACCCAAACCACCCGTATATGAACCTCCAAAGAAGGAGAAGCCCGAGCCAAGGCCAC
421 480
P E P K P P V Y E P P K K E K P E P R P
CGGTTTATACACCTCCAAAGAAAGAGAAACCTGAACCCAAACCACCAGTGTATGAACCTC
481 540
P V Y T P P K K E K P E P K P P V Y E P
CAAAAAAACCTCCAATGTACGAACCTAAGCCACCAAAACCACCGGTTTATACACCTCCAA
541 600
P K K P P M Y E P K P P K P P V Y T P P
AGAAAGAGAAACCTGAACCCAAACCACCAATGTATGAACCTCCAAAAAAACCTCCAATGT
601 660
K K E K P E P K P P M Y E P P K K P P M
ATGAACCTAAGCCACCGAAGCCACCGGTTTATACACCTCCAAAGAAAGAAAAACCAGAAC
661 720
Y E P K P P K P P V Y T P P K K E K P E
CCAAGCCACCAATGTATCAACCTCCCAACAACCCACCAATATATGAGCCAAAACCACCTA
721 780
P K P P M Y Q P P N N P P I Y E P K P P
AGCCACCAGTTTATGCACCACCAAAGGAAGAAAAGCCAAAACCTAAACCACCAGTTTATG
781 840
K P P V Y A P P K E E K P K P K P P V Y


ACGCTCCGGCACATGAGCCACCATACGGTCACTATCCAGGACATCCACCTTTGGGGAAGC
841 ---------+---------+---------+---------+---------+---------+ 900
D A P A H E P P Y G H Y P G H P P L G K
pFS14
CTCAATAGAATGCTCCCATCGACTAGCCATTCCCAACAATTGAGCTACTAGAAAAATTAG
901 ---------+---------+---------+---------+---------+---------+ 960
P Q *
TATTTATAATAAGAGGAACTGCCGGCTGATGTTCTGTTTTATTCTTGTATGGAATGTAGT
961 ---------+---------+---------+---------+---------+---------+ 1020
pFS17
P17.2
P17.1
AATGTTTTCTTTCATGTTTGAAGAGAGACCAGTTGCTCATGGCTGTACTCATTCGAATGC
1021 ---------+---------+---------+---------+---------+---------+ 1080
TTATTCTATATCTTCTTCTTTCTCAATGAGCAATATCCTGATGATTTCAAATAAAAAAAA
1081 ---------+---------+---------+---------+---------+---------+ 1140
AAAAAAAAAAAAAAA
1141 ---------+----- 1160

et al., 1989; Datta and Marcus, 1990; Hong *et al.*, 1990). The repetitive nature of PRPs renders sequence comparison difficult, as multiple alignments can be constructed for each pairwise sequence comparison. Nevertheless, the highest degree of sequence similarity was found to a soybean cDNA coding for a repetitive proline-rich protein SbPRP1 (Hong *et al.*, 1987) with 57% nucleotide identity over a 385 bp overlap. At the protein level, the preFSPRP and SbPRP1 sequences were 47% identical over a 154 amino acid overlap.

4.5.3e Proline-rich Proteins in Plant Cells

PRPs (Section 1.4.2a) represent a class of cell wall proteins, the members of which have a preponderance of hydroxyproline residues. They are characterised by the occurrence of Pro-Pro repeats that are contained within a variety of larger repeat units, generally pentapeptide units of the form Pro-Pro-X-Y-Lys, where X and Y can be valine, lysine, tyrosine, histidine and glutamic acid (Keller, 1993; Showalter, 1993). Some PRPs (such as SbPRP1, SbPRP2 and SbPRP3 from soybean) are composed only of tandem repeats of the above form whereas others, such as carrot p33 (Chen and Varner, 1985) and PvPRP1 from bean (Sheng *et al.*, 1991), contain unique regions of sequence interspersed with the repetitive elements.

The preFSPRP sequence displays many of the features common to plant cell wall PRPs. Analysis of the predicted amino acid sequence revealed an N-terminal region with several of the features characteristic of eukaryote signal peptides (Nothwehr and Gordon, 1990) and the predicted signal sequence cleavage site of preFSPRP was calculated according to von Heijne (1983) and confirmed by alignment of the N-terminal sequence of FSPRP with those of PRPs present in other dicotyledonous plants (Figure 4.18a). Presence of the signal sequence, together with absence of an endoplasmic reticulum signal KDEL or HDEL (Denecke *et al.*, 1992) suggests that the preFSPRP mature protein, like other PRPs, is targeted to the secretory pathway in fibre cells and thus localised in the extracellular matrix.

Hydropathy profiles (Figure 4.18b) of the preFSPRP amino acid sequence showed that

a.

FSPRP	**T Y E S P N Y G K P P T P V F K P P K**	
SbPRP1	D **Y E** K **P** P I Y **K P P** V Y T **P P V Y K**	(soybean; Hong *et al.*, 1987)
SbPRP2	N **Y E** N **P** P V Y **K P P T** E K **P P V Y K**	(soybean; Hong *et al.*, 1990)
SbPRP3	**T Y** Y K - P I K **K P P** V Y K **P P V Y K**	(soybean; Hong *et al.*, 1990)
MtPRP4	N **Y E** K **P** P E Y **K P P** V E N P Q F Y **K**	(alfalfa; Wilson *et al.*, 1994)
DcPRP1	S H S H **P** P I H **K P P** V Y T P P V H **K**	(carrot; Ebener *et al.*, 1993)

b.

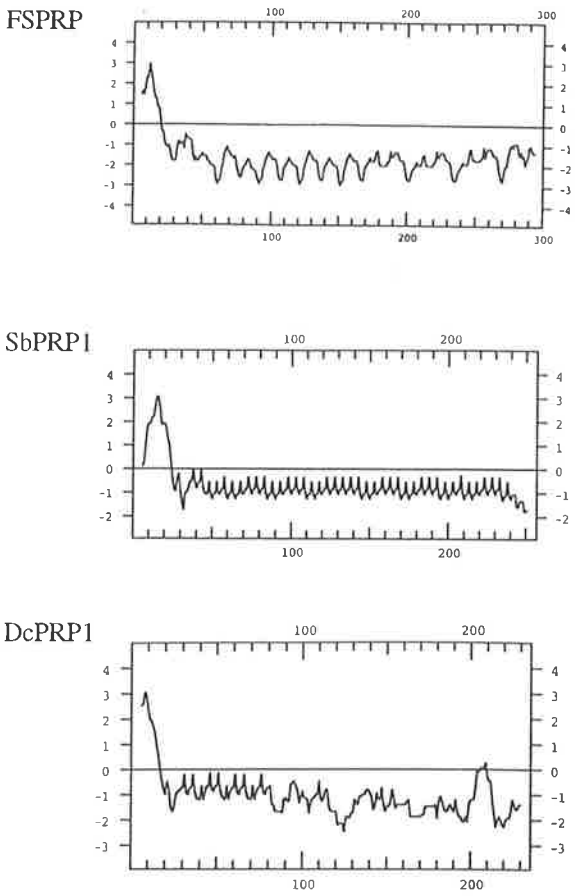


Figure 4.18: A cotton fibre-specific PRP

a. N-terminal sequence alignment of the mature cotton fibre PRP, FSPRP, with N-terminal sequences found in PRPs of other dicotyledonous plants. Amino acids are indicated by single-letter nomenclature and the amino acids which are in bold type indicate identity with the FSPRP sequence. The first repeat unit in the three soybean PRPs is underlined.

b. Comparison of the hydropathy plots of a cotton fibre-specific PRP with the homologues from soybean (SbPRP1; Hong *et al.*, 1987) and carrot (DcPRP1; Ebener *et al.*, 1993). The horizontal scale indicates the number of amino acid residues and the vertical one the relative scale, according to Kyte and Doolittle (1982). Points above the horizontal line correspond to hydrophobic region and points below this line are hydrophilic. SbPRP1 contains 43 repeats of a sequence consisting primarily of PPVYK, whereas the less well-defined repetitive elements in DcPRP1 are interspersed with unique regions of sequence.

the hydrophobic signal sequence is followed by a highly hydrophilic region which is representative of the mature protein. Assuming that the cleavage site of the protein is between amino acid residues 23 and 24 (Figures 4.17 and 4.18a), the mature protein, designated FSPRP, has 276 amino acids with a calculated molecular weight of 31567 Da and a pI of 10.5, which fall into the range reported for other plant cell wall PRPs. Also apparent from the hydropathy profile is the repetitive nature of the protein sequence, which is typical of plant PRPs (Figure 4.18b) but which can diverge widely between PRPs from different plants.

PRPs show a large degree of heterogeneity in their amino acid compositions and PRPs may even tolerate the loss of integral repeat units (Showalter and Rumeau, 1990; Keller, 1993; Showalter, 1993; Schmidt *et al.*, 1994). The FSPRP protein is highly enriched in proline (111 residues, 40%), lysine (59 residues, 21%), glutamic acid (33 residues, 12%) and tyrosine (23 residues, 8%), and therefore resembles those PRPs isolated from soybean, which lack histidine and serine but have high levels of tyrosine and lysine. FSPRP has a greater amount of acidic amino acids (such as Glu) than the majority of soybean PRPs. Particularly interesting in the FSPRP deduced amino acid sequence is its repeat structure. The FSPRP repeat unit appears to be related to the plant PRP motif PPVYK (Showalter, 1993). Josè and Puigdomènech (1993) classified PRPs into four groups based on the repeat unit sequence, and the sequences may diverge widely from the general element PPVYK. With the aid of the hydropathy profile (Figure 4.18b), it is possible to divide the preFSPRP protein into four regions based on repeat sequence structure (Figure 4.19). The first includes the hydrophobic signal peptide and the 28 N-terminal residues of the mature protein, both of which have little or no repeat sequence (Figure 4.19a). The centre region of the protein consists of two distinct parts. The first (Figure 4.19b) contains small repeats of 15 residues each, within which can be identified regions similar to the general element PPVYK (Josè and Puigdomènech, 1993). In a similar way, the repetitive pentapeptides in a soybean PRP (Hong *et al.*, 1990) and in the homologue from alfalfa (Wilson and Cooper, 1994) are arranged into longer decapeptide repeats. The second repetitive region

a. MMSTTHLLVFLLGVVTLTTPTFG↑TYESPNYGKPPTPVFKPPKVKPPPYEPK

b. PPVYEPPKKEKPEPK
PPVYAPPKKEKHGPK
PPMYEPPKKEKPEPK
PPVYTPPKKEEPKPK
PPVYEPPKKEKPEPK
PPIYTPPKKEKPEPK
PPVYEPPKKEKPEPR
PPVYTPPKKEKPEPK
PPVYE

c. PPKKPPMYEPKPPKPPVYTPPKKEKPEPKPPMYE
PPKKPPMYEPKPPKPPVYTPPKKEKPEPKPPMYQ
PPNNPPIYEPKPPKPPVYAPPKEEKPKPKPPVYD

d. APAHEPPYGHYPGHPPLGKPQ*

Figure 4.19: The repeat structure of a cotton fibre-specific PRP

The amino acid sequence of a cotton fibre-specific PRP, divided into four regions or domains, based on their apparent repeat unit structure and the hydropathy profile of preFSPRP. Sequences reminiscent of the general PRP element, PPVYK, are underlined.

a. Amino acids 1 to 51 of the preFSPRP conceptual amino acid sequence, lacking in repeat structure. The putative signal peptide cleavage site is marked with an arrow.

b. Amino acids 52 to 176 of the preFSPRP conceptual amino acid sequence, arranged in tandem repeats of 15 amino acids each. Included is the intervening region between the blocks of repeats, consisting of five amino acids.

c. Amino acids 177 to 278 of the preFSPRP conceptual amino acid sequence, arranged in tandem repeats of 34 amino acids each.

d. Amino acids 279 to 299 of the preFSPRP conceptual amino acid sequence, lacking in repeat structure.

consists of three direct tandem repeats of 34 residues each and of the sequence PP(K/N)(K/N)P P(M/I)YEP KPPKP PVY(T/A)P PK(K/E)EK P(E/K)PKP P(M/V)Y(Q/D/E) (Figure 4.19c). The carboxy terminus of the FSPRP protein, like the amino terminus, is apparently devoid of a repeat structure (Figure 4.19d).

PRPs are encoded by a small multigene family in several plant species, the members of which are tightly regulated and exhibit cell- and tissue-specific patterns of expression as well as developmental regulation (Keller, 1993; Showalter, 1993; Schmidt *et al.*, 1994). PRPs are best characterised in soybean, for which several different cDNAs (Hong *et al.*, 1987; Averyhard-Fullard *et al.*, 1988; Datta *et al.*, 1989) and three different genomic clones (Hong *et al.*, 1987; Datta and Marcus, 1990; Hong *et al.*, 1990) have been isolated. Individual PRP mRNAs are expressed differentially in the apical, elongating and mature regions of the soybean hypocotyl (Wyatt *et al.*, 1992).

Regulatory studies indicate that members of the PRP gene family from several plants are involved in various aspects of development, ranging from germination to the early stages of nodulation, and expression is often limited to specific cell types or organs (reviewed in Showalter and Rumeau, 1990). In addition to their developmental regulation, several PRP cell wall proteins have been implicated in plant defense reactions and as a response to wounding. For example, insolubilisation of PRPs has been observed in wounded bean hypocotyls and after fungal elicitor treatment of soybean cell cultures (Bradley *et al.*, 1992). PRP expression in carrot (Chen and Varner, 1985; Ebener *et al.*, 1993), bean (Sheng *et al.*, 1991) and soybean (Kleis-San Francisco and Tierney, 1990) is wound inducible. Immunolocalisation of cell wall PRPs in several plant species seems to indicate that PRP localisation is related to the pattern of lignification (Ye *et al.*, 1991), however some of the cells in which soybean PRP gene expression is found have only primary walls which are not lignified (Wyatt *et al.*, 1992). Moreover, work in this thesis has isolated a PRP which is abundant in cotton fibres, a cell type which lacks lignin. These data, together with the complex patterns of expression observed for

members of PRP gene families, preclude a simple correlation of PRP function with a known cell wall structure. It is possible that PRPs could play different roles in different cell types.

The precise function of PRPs is therefore unknown, but a general role in the strengthening of plant cell walls is envisaged. PRPs are not solubilised by high salt (Datta *et al.*, 1989) and are thought to be bound to the cell wall in complex structures. They may form a cross-linking network, as proposed for extensins (Fry, 1986; Cooper *et al.*, 1987), or they may interact with other components of the cell wall. Support for the latter hypothesis comes from the demonstration that many proline-rich proteins isolated from bacterial and mammalian systems are involved in binding processes which are generally mediated by the proline residues (Williamson, 1994). PRPs, typically with tandemly repeated sequences 5-8 residues in length, have been demonstrated to bind cytoskeletal proteins such as actin. In addition, mammalian salivary and some fungal PRPs probably function to bind plant polyphenols (Murray *et al.*, 1994). Such a cross-linking network would strengthen the cell wall against environmental stress and pathogen attack (Bradley *et al.*, 1992).

4.5.3f A PRP Specific to Cotton Fibres

Proteins such as PRPs, which presumably have a structural role in plant cell walls, are expected to be abundant in rapidly elongating cells such as cotton fibres, since such growth requires a continuous supply of wall materials. An increase in mRNA levels of a barley PRP has been observed in cells of the leaf extension zone of a *slender* mutant (Schünmann *et al.*, 1994). The *slender* phenotype is characterised by an increased leaf extension rate, resulting in part from epidermal cells which are both longer and narrower than wild type. Expression of several PRP genes is responsive to auxin, a plant growth factor which is involved in the enhancement of cell elongation and enlargement (Hong *et al.*, 1987; Ebener *et al.*, 1993) and has been implicated in control of cotton fibre development (Section 1.3.3c). It is notable that FSPRP transcripts are abundant early in fibre development, during elongation, and that mRNA

levels decrease at the onset of secondary wall synthesis when mainly cellulose is deposited (Figure 4.4b). However, it is unknown whether the levels of FSPRP correlate with the abundance of its mRNA and previous work has identified the possibility of post-translational gene regulation in cotton fibres (John and Crow, 1992; John and Keller, 1995). Interestingly, SbPRP1 in soybean is predominantly expressed in epidermal cells and in the seed coat (Wyatt *et al.*, 1992). It would be of interest to determine whether FSPRP expression occurs in all cells of the seed coat or whether expression is confined strictly to fibre cells.

The nucleotide sequence of cFS17 and the deduced amino acid sequence did not show significant similarity to a proline-rich protein previously isolated from cotton fibres, H6 (John and Keller, 1995). The H6 gene, like pFS17, is expressed predominantly in cotton fibre cells and the transcript is present during early primary cell wall formation. Cotton protein H6 contains 17 copies of a repetitive pentameric motif of (Ala/Ser)(Thr/Ser)Pro Pro Pro which is clearly different from the FSPRP repeat unit(s) and places H6 in the arabinogalactan-binding group of proteins or AGPs (Section 1.4.2a). The appearance of the H6 protein coincides with the onset of secondary wall synthesis and suggests a translational regulation of H6 expression, as well as transcriptional regulation, and a possible role for the protein in the assembly of the cell wall matrix.

Yet another fibre-specific protein, B6 (John, 1995), contains two proline-rich peptide regions, both of which are dissimilar to the repeats identified in the FSPRP sequence. The B6 mRNA is present during both the primary and secondary cell wall synthesis stages of fibre development and the encoded protein contains a hydrophobic N-terminus which may be indicative of presence of a signal peptide. The function of the B6 protein in fibre cells is unknown.

4.5.3g Isolation of a Putative pFS17 Genomic Clone

The insert fragment from pFS17 was used to screen a partial genomic library consisting

of 1.6 x 10^5 recombinants, resulting in a single isolate which only weakly hybridised the probe. The positive clone was previously identified in the screening of the genomic library with the pGhEF-1α insert (clone λS5) and its characterisation is described in Section 3.6. Both probes showed weak hybridisation to the genomic clone λS5 and pFS17, like pGhEF-1α, hybridised to the *Xho*I and *Sac*I doublet fragments of λS5 (Figure 3.9). pFS17 also hybridised a 3.1 kb *Xho*I fragment which is attached to the short arm of λGEM-11™ in clone λS5 (Figure 3.9), together with the corresponding 12 kb *Eco*RI fragment and 2.8 kb *Sac*I fragment. Since pFS17 hybridises more strongly to both these fragments than does pGhEF-1α, it may be that the 2.8 kb *Sac*I fragment of λS5 contains sequences unique to the pFS17 probe, or regions that are more similar to the pFS17 sequence (Figure 4.13a) than to the pGhEF-1α sequence (Figure 3.9).

It is possible that λS5 contains the gene which corresponds to the fibre-specific pFS17 cDNA. The repeat unit of approximately 4.5 kb is sufficient in size to contain the coding regions for the 1.2 kb pFS17 cDNA, in addition to possible coding regions for an EF-1α gene. However, given the relatively weak hybridisation of pFS17 to the genomic clone, and the equal hybridisation of an apparently unrelated sequence, λS5 is unlikely to contain the FSPRP gene. Further characterisation of genomic clone λS5 was not pursued beyond the work outlined in Chapter 3. Probing of the genomic library with the complete 1.2 kb cDNA may result in isolation of further positives which hybridise more strongly to probe sequences and are therefore better candidates for the fibre-specific PRP gene.

4.5.3h Southern Analysis of Fibre-specific Clone pFS17

Presence of a PRP gene family in cotton was confirmed by Southern analysis. Probing a Southern blot of restricted cotton genomic DNA with the pFS17 insert resulted in two to five hybridising genomic fragments for each enzyme (Figure 4.20). The cotton genome contains a small PRP gene family, the members of which have varying degrees of similarity to the pFS17

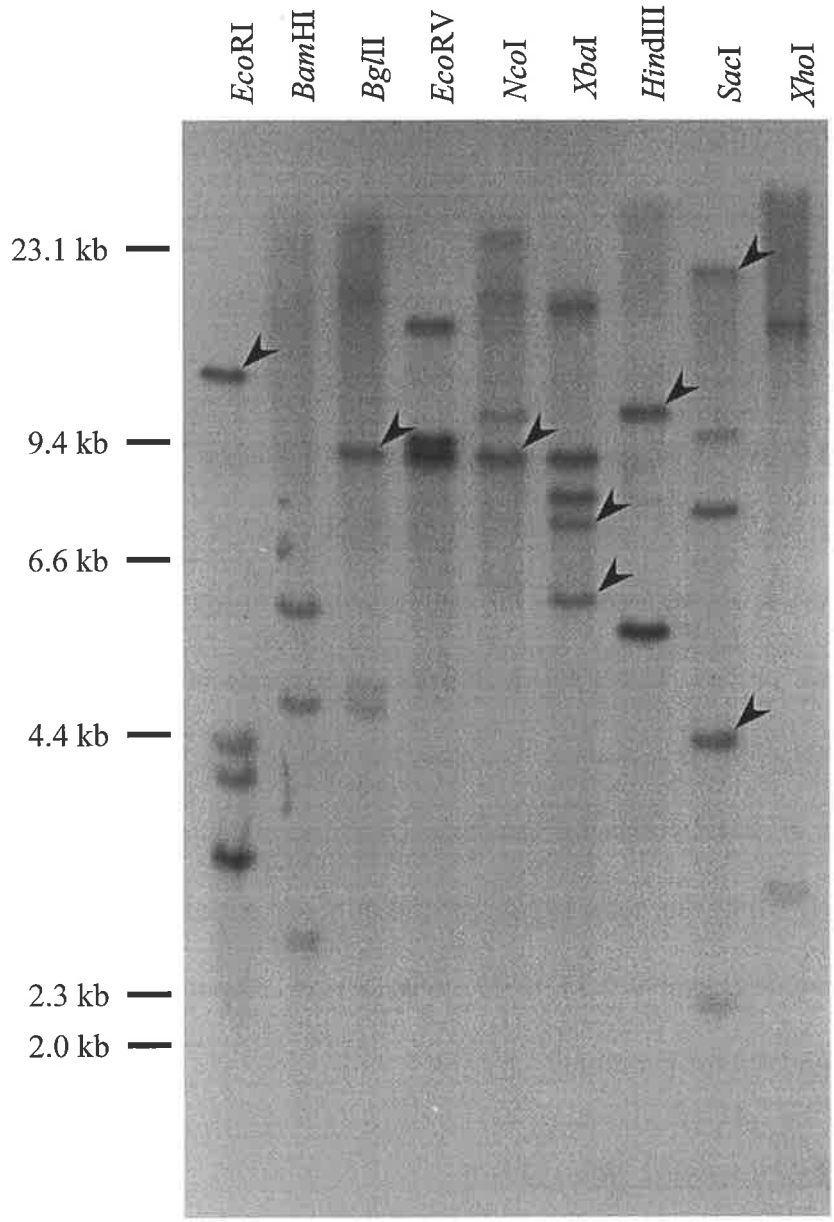


Figure 4.20: Southern analysis of fibre-specific cDNA clone pFS17

Genomic Southern blot probed with the pFS17 insert. Each lane was loaded with 10 µg of total genomic DNA cut with restriction enzymes which do not contain recognition sites within the cFS17 sequence. Numbers on the left represent the molecular weights of wild type Lambda DNA restricted with *Hin*dIII (kb) and arrows indicate genomic fragments of a similar size as those hybridised by pGhEF-1α. Lanes are *G. hirsutum* L., cv. Siokra 1-2 DNA restricted with endonucleases as indicated.

cDNA clone. Observation of low-level hybridisation to mRNAs in cotton tissues other than fibres (Figure 4.4b) implies that pFS17 may be detecting transcripts from another member of the PRP gene family, depicted in Figure 4.20.

The pFS17 Southern, as expected, contained some hybridising bands which were similar in size to those obtained when an identical Southern was probed with pGhEF-1α (Figure 3.7; arrowed in Figure 4.20). Restriction of genomic DNA with *Eco*RI resulted in a commonly hybridising fragment of around 12 kb in size, and restriction with *Bgl*II (a common band of approximately 9 kb; not shown in Figure 3.7), *Nco*I (a common band of approximately 9 kb; not shown in Figure 3.7), *Xba*I (common bands of approximately 6 kb and 7 kb), *Hin*dIII (a common band of approximately 10 kb) and *Sac*I (common bands of approximately 4 kb and 18 kb) also resulted in genomic fragments which hybridised both probes. The 4 kb *Sac*I band corresponds in size to the 4.5 kb restriction fragment (doublet) of S5 and most likely corresponds to this fragment of the genomic clone. The majority of the hybridising genomic DNA fragments were unique to each clone, suggesting that, whilst there are sequences common to both pFS17 and pGhEF-1α contained in a single genomic fragment, λS5, there are sequences unique to each clone which are widely dispersed throughout the genome.

4.5.4 Cotton Fibre-specific clone pFS18

4.5.4a Initial Sequence Characterisation of Fibre-specific Clone pFS18

pFS18, like pFS3, did not detect similar sequences in any other fibre-specific cDNA clones (Figure 4.3b). The insert of pFS18 was fully sequenced and at 127 bp, was among the shortest of the fibre-specific cDNA clones (Figure 4.21). The sequence was AT-rich, with 76% AT, and a sequence identical to the polyadenylation consensus sequence 5'-AATAAA-3' was identified at nucleotide position 76 in the pFS18 sequence.

Database searches with the pFS18 sequence failed to identify any closely related genes or cDNAs (GenBank and EMBL, September 1994). This is not unexpected given the length of

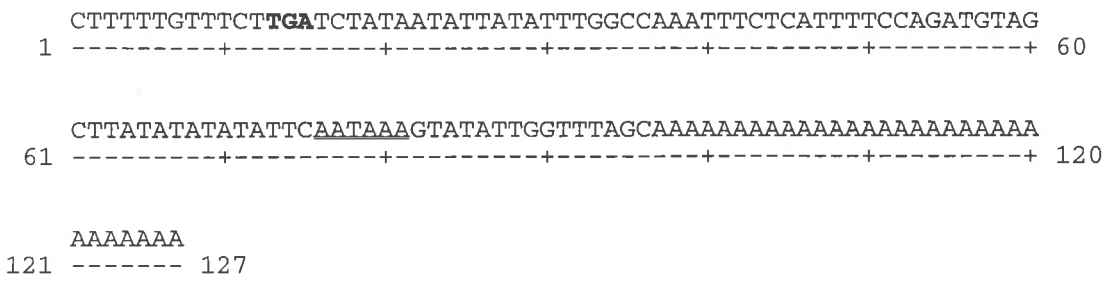


Figure 4.21: Nucleotide sequence of cDNA clone pFS18

Complete insert sequence of fibre-specific cDNA clone pFS18. The putative polyadenylation signal is double underlined and a possible stop codon TGA is in bold.

the sequence and the possibility that, while a putative TGA stop codon is present in the pFS18 sequence (Figure 4.21), the clone may not contain any coding sequences. pFS18 may represent only the 3' untranslated region of the corresponding fibre-specific mRNA. The sequence of pFS18 diverged from fibre-specific cDNAs isolated from other cotton varieties (John and Crow, 1992; Delmer *et al.*, 1995; John, 1995; John and Keller, 1995; Ma *et al.* 1995) suggesting that pFS18 is a novel fibre-specific cDNA. Despite the fact that pFS18 was clearly fibre-specific in the differential screen (Figure 4.2), it was an ineffective probe to Northern blots of cotton RNAs.

The AT-rich nature of the sequence rendered pFS18 unsuitable for 5'-RACE experiments, since 5'-RACE requires two primers of at least 25 bp each, both with GC contents of 45-65%. The sequence of pFS18 was too short to allow design of primers which did not undergo extensive intra- and inter-molecular hydrogen bonding. In addition, primers complementary to pFS18 would have a low binding specificity due to their low GC content and would most likely amplify the 5' ends of cDNAs other than pFS18.

In an effort to obtain a full-length cDNA clone, longer pFS18 homologues were isolated from the 13 DPA fibre cDNA library. One homologue was used to demonstrate the fibre-specific expression of pFS18, and its temporal expression through fibre development (Figure 4.4b) was determined. pFS18 transcripts peak in concentration at around 12-14 DPA, at a time immediately prior to onset of the secondary wall synthesis phase of fibre growth. It is unknown whether the appearance of the pFS18 protein correlates with its accumulated transcript, but it is nevertheless tempting to speculate that the pFS18 protein has a role in secondary wall synthesis such as the one proposed for a fibre-specific GTP-binding protein (Delmer *et al.*, 1995; Section 1.6.2).

4.5.4b Isolation of a Putative pFS18 Genomic Clone

In order to isolate genomic sequences corresponding to cDNA clone pFS18, a Siokra

1-2 partial genomic library consisting of 1.6 x 10^5 clones was screened with the pFS18 insert. Three clones showed weak hybridisation to the probe, one of which was the pGhEF-1α genomic clone λS5 (Chapter 3). Since two other cDNA clones, pGhEF-1α and pFS17 also show weak homology to λS5, the observed weak hybridisation by pFS18 to λS5 can be considered to be non-specific. The three cDNA clones which hybridise λS5 do not show any significant pairwise sequence similarity other than that they are all AT-rich.

Of the remaining two positives, λFS18(A) was purified and partially restriction mapped (Figure 4.22). It was apparent from the restriction patterns that clone λFS18(A) had an unusual structure in that the left arm of the λGEM-11™ vector appeared to contain a single internal *Bam*HI site. Complete digests using combinations of the three enzymes failed to resolve the map. The 4.6 kb *Bam*HI fragment was uncut in double digests with *Sac*I and *Xho*I, and oligonucleotide ON-L annealed to a fragment of approximately 4.6 kb in a complete *Bam*HI restriction of λFS18(A) DNA (not shown). Bacteriophage vectors such as wild type lambda, λZAP®II (Short *et al.*, 1988) and λgt10 (Huynh *et al.*, 1985) possess *Bam*HI sites at nucleotide position 5510 in the sequence of their left arms, but the genomic cloning vector used in this study, λGEM-11™, is a derivative of EMBL3 (Frischauf *et al.*, 1983) and contains two *Bam*HI sites, which are present in the duplicated multiple cloning site. In addition, restriction with *Sac*I did not produce the expected 9 kb derivative of the right arm of λGEM-11™, although a 20 kb restriction fragment corresponding to the left arm of the vector was produced (not shown). Similarly, restriction of λFS18(A) DNA with *Xho*I did not generate either of the vector arms. The only conclusion which can be drawn is that the λGEM-11™ used in construction of the genomic library contains a small amount of contaminant arms from a phage other than λGEM-11™ and it is into these that the 14.2 kb λFS18(A) insert was cloned.

Southern experiments showed that the insert from pFS18 had the greatest amount of

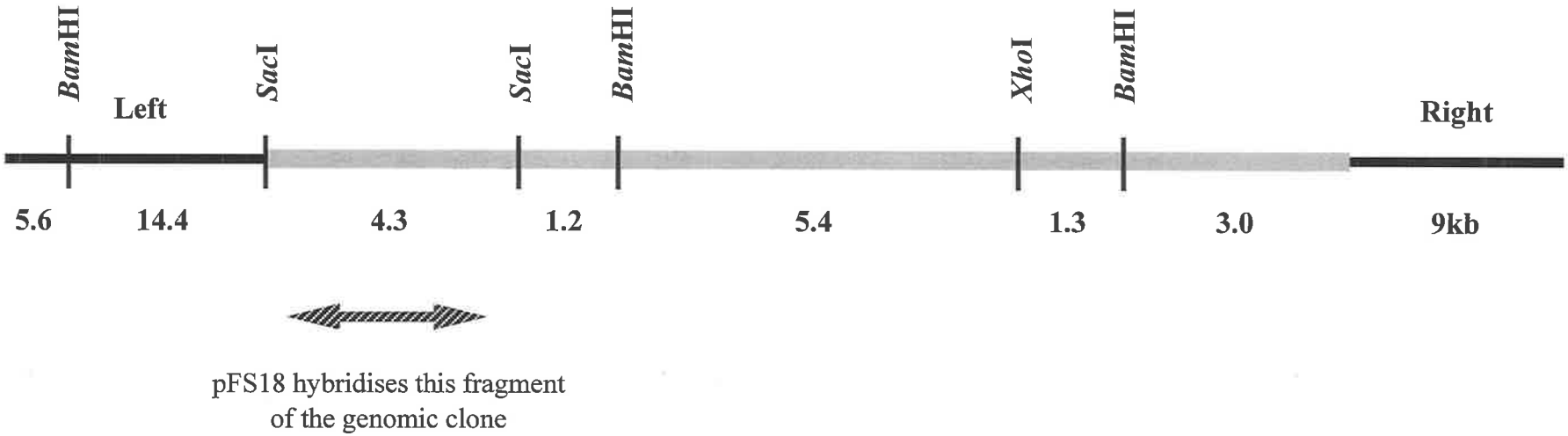


Figure 4.22: Partial restriction map of genomic clone λFS18(A)

Restriction map of genomic clone λFS18(A), indicating the positions of the *Xho*I, *Bam*HI and *Sac*I sites, with fragment sizes given in kb and the region hybridised by pFS18 designated by an arrow.

homology to sequences in the 4.5 kb *Sac*I fragment of λFS18(A) (Figure 4.22). pFS18 also hybridised weakly to λS5 DNA, showing approximately equal hybridisation to each of the three *Sac*I and *Xho*I insert restriction fragments (not shown). The degree of hybridisation by pFS18 was similar to each of the genomic clones (not shown), and the insert from clone pGhEF-1α (Chapter 3) showed a comparable degree of hybridisation to λS5 and λFS18 sequences (not shown).

Attempts to subclone the 4.5 kb *Sac*I fragment of λFS18(A) into a plasmid vector were unsuccessful and since probe pFS18 showed a similar degree of hybridisation to each of the genomic clones λFS18(A) and λS5, the characterisation of λFS18(A) was not pursued further. Use of longer pFS18 cDNA homologues to rescreen the genomic library may result in the isolation of further genomic clones which are better candidates for the pFS18 gene than λFS18(A) and λS5.

4.6 Summary and Conclusions

Five fibre-specific cDNA clones were isolated by differential screening of a cDNA library from cotton fibres, a developmentally synchronous population of non-dividing cells. The genes corresponding to these cDNAs were expressed preferentially in fibre cells and exhibited differing patterns of temporal expression during fibre development. All of the cDNA clones correspond to abundant transcripts in fibre cells 13 days after anthesis. Observation of low-level hybridisation to mRNAs in cotton tissues other than fibres may indicate a basal level of expression in all tissues, or cross-hybridisation to transcripts from other diverged members of the gene family, as may be the case for pFS1, pFS3 and pFS17.

Conceptual translations of three of the clones showed significant amino acid sequence similarity to previously characterised plant proteins and all three detected transcripts which were most abundant in the primary wall synthesis, or elongation, phase of fibre development.

The function of pFS1, which showed homology to a previously reported fibre-specific sequence (John and Crow, 1992), is unknown but the protein is thought to play a structural or enzymatic role in fibre development.

Appearance of fibre-specific proline-rich protein (PRP) transcripts during elongation of the fibre cell is consistent with a presumed structural role for FSPRP in the strengthening of the cell wall. Such an event would occur during primary wall synthesis, as secondary wall synthesis in cotton fibres consists primarily of cellulose deposition. It should be emphasised that the FSPRP cDNA has not been shown to encode a cell wall protein. However, circumstantial evidence strongly suggests that the pFS17 cDNA encodes a protein destined for the cell wall as its conceptual translation shows numerous properties similar to PRPs from other higher plants. Based on sequencing studies of large numbers of fibre-specific cDNA clones, John and Keller (1995) suggest that H6, an AGP-like fibre-specific protein, is the major proline-rich protein in fibre. However, our results suggest that in *G. hirsutum* L., cv. Siokra 1-2, FSPRP is the major proline-rich protein in the walls of cotton fibres.

Sequence analysis of the remaining two clones suggests that they may encode as yet uncharacterised protein products. Since pFS3 transcripts are detectable in fibres aged 6 DPA to 20 DPA, it is conceivable that the protein has a general role in fibre development and may be involved in both primary wall synthesis and secondary wall deposition. The abundance of pFS18 mRNA appeared to peak at 12 to 14 DPA, prior to the time of transition between the elongation (primary wall synthesis) and secondary wall deposition phases of fibre growth, suggestive of a function for the pFS18 protein in secondary wall deposition and possibly in the deposition of cellulose microfibrils.

In a similar study, John and Crow (1992) did not detect any subset of genes that are expressed during a given developmental stage, prompting the suggestion that most of the genes in cotton fibres are active throughout the development of the fibre cell. In contrast, we have isolated only one fibre-specific cDNA, pFS3, whose transcripts are present in abundance

throughout the development of the fibre, up to 20 DPA. Transcripts corresponding to the remaining four genes vary in abundance as the fibre develops, suggestive of temporal control of gene transcription in addition to spatial control. Results obtained in these experiments thus indicate a role for transcriptional control of gene expression in cotton fibre cells, in addition to the translational control observed for other fibre-specific genes (John and Crow, 1992; John and Keller, 1995).

CHAPTER 5

Selective Expression of a Lipid Transfer Protein Gene in Developing Cotton Fibre Cells

5.1 INTRODUCTION

Five fibre-specific cDNA clones were isolated by differential screening of a cDNA library from cotton fibres and their partial characterisation is described in the previous Chapter. The fibre-specific transcripts were abundant in fibre cells and exhibited differing patterns of temporal expression during fibre development, but were essentially absent from other cotton tissues. Transcripts corresponding to one clone, pFS6, represented the most plentiful class of fibre-specific cDNAs and is the most abundant fibre-specific mRNA in 13 DPA cotton fibres (Table 4.1).

Described in this Chapter is the characterisation of fibre-specific cDNA clones within the pFS6 group, together with the isolation and characterisation of the full-length pFS6 cDNA and the partial characterisation of two corresponding genes from the Siokra 1-2 genome.

5.2 SEQUENCE ANALYSIS OF PFS6-RELATED CDNA CLONES

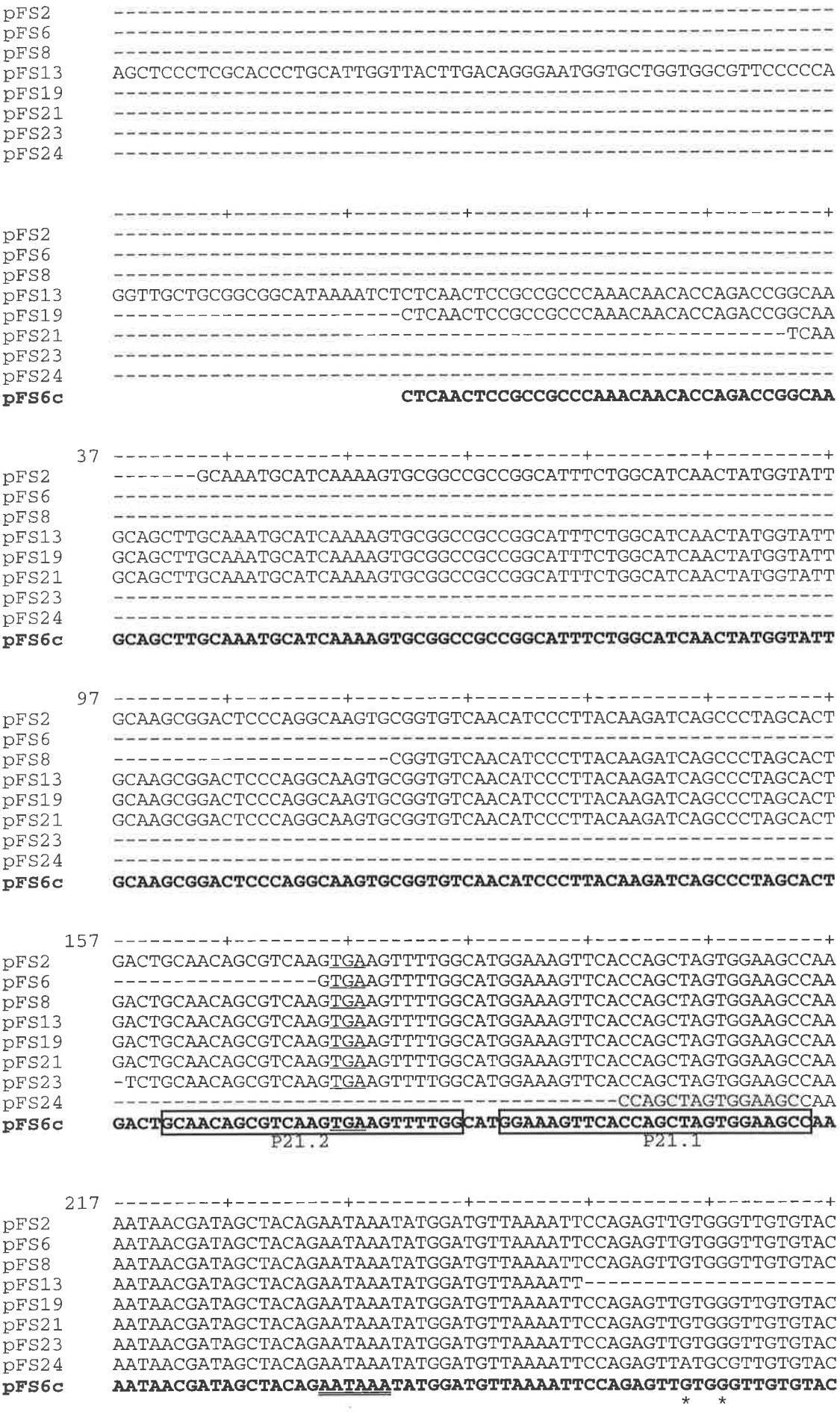
Sequence information for cDNA clone pFS6 was used in an initial attempt to determine potential function of the encoded protein. However, screening of the nucleotide databases (GenBank and EMBL, December 1994) with the 149 bp sequence of pFS6 was uninformative. The short length of the pFS6 sequence and presence of a putative termination codon TGA at nucleotide position two implied that the clone did not contain any coding sequences and consisted of 3' UTR only.

Further pFS6-like cDNA clones were sequenced in order to obtain a pFS6 sequence which extended into the coding region of the pFS6 mRNA. Twelve pFS6-related fibre-specific cDNA clones were identified in the differential screen, ranging in size from approximately 100 bp (pFS9) to 800 bp (pFS13) (Figure 4.3). Eight of the clones, pFS2, pFS6, pFS8, pFS13, pFS19, pFS21, pFS23 and pFS24, were sequenced, the results of which are summarised in Table 5.1 and presented as an alignment in Figure 5.1. The inserts of all pFS6 homologues depicted in Figure 5.1 except pFS13 and pFS23 were sequenced to completion.

Figure 5.1: Sequencing of the pFS6 group clones

Alignment of nucleotide sequences of eight pFS6 group cDNA clones, not including poly(A) tails, together with a consensus sequence, pFS6c. Numbering refers to nucleotides in this consensus sequence. Putative TGA stop codons in each sequence are underlined and the regions complementary to the 5'-RACE cDNA primer (P21.1) and the 5'-RACE PCR primer (P21.2) are boxed. Sequence differences between the clones are marked with asterisks below the consensus sequence and putative polyadenylation signals are double underlined.

Of the clones with multiple inserts, pFS13, pFS23 and pFS24, only the relevant sequence is shown in the alignment. The partial sequence of cDNA clone pFS13 from the RSP of pBluescript®SK(-) is shown, and extends from nucleotide 1 to nucleotide 255 of the pFS6c sequence. The additional 5' sequence of pFS13R was not incorporated into the pFS6c consensus sequence since the possibility of a multiple insert in this region could not be excluded.


pFS2
pFS6
pFS8
pFS13
pFS19
pFS21
pFS23
pFS24
pFS6c
P21.2
P21.1

```
     277 ---------+---------+---------+---------+---------+---------+ 329
pFS2     TATGCCGCTTTATGCGACTACGTAATATTAACTTTATCTACAAATTAATATCA-------
pFS6     TATGCCGCTTTATGCGACTACGTAATATTAACTTTATCTA-AAATTA-------------
pFS8     TATGCCGCTTTATGCGACTACGTAATATTAACTTTATCTACAAATTAATATCACTCGTCT
pFS13    ------------------------------------------------------------
pFS19    TATGCCGCTTTATGCGACTACGTAATATTAACTTTATCTACAAATTAATATCACTCGTCT
pFS21    TATGCCGCTTTATGCGACTACGTAATATTAACTTTATCTACAAATTAATATCA-------
pFS23    TATGCCGCTTTATGCGACTACGTAATATTAACTTTATCTACAAATTAATATCA-------
pFS24    TATGCCGCTTTATGCGACTACGTAATATAATCTTTATCTACAAATTAGTATCA-------
pFS6c    TATGCCGCTTTATGCGACTACGTAATATTAACTTTATCTACAAATTAATATCA
                                  * *         *       *

pFS2     ----------
pFS6     ----------
pFS8     CCATTTCCCA
pFS13    ----------
pFS19    CCATTTCCCA
pFS21    ----------
pFS23    ----------
pFS24    ----------
```

Table 5.1 Sequencing of the pFS6 group clones

Category of transcript length	pFS6 clone group	Fibre-specific cDNA clone	Insert size (bp)	Poly(A) tail	Features of note
Short	Group I	pFS6	149	$(A)_{39}$	Defining and shortest clone, has a single nucleotide deletion relative to the Group II sequence
Medium	Group II	pFS2	286	$(A)_{18}$	Longer than pFS6 by six 3'-terminal nucleotides
		pFS21	296	$(A)_{24}$	Identical to pFS2 but extended 10 bp further into the coding sequence
		pFS23	171	$(A)_{7}$	Contains multiple inserts but 171 bp is the limit of identity to pFS2
	Group III	pFS24	131	$(A)_{72}$	Identical to the Group II clones in length but contains five nucleotide substitutions; comprises multiple inserts
Long	Group IV	pFS8	227	$(A)_{66}$	Identical to the Group II sequence except for an additional 17 bp to the 3' end
		pFS19	346	-	Identical in sequence to pFS8 but lacks a poly(A) tail
Unknown	Unknown	pFS13	339	-	Incomplete at the 3' end but otherwise identical to the Group II sequence; comprises multiple inserts

Sequence alignment of the eight clones revealed variation in both sequence and length at their 3' ends (Figure 5.1) and enabled the clones to be divided into four groups (Table 5.1). Three length variants were observed in addition to some minor sequence differences between the clones. An intermediate class of transcript encompasses clone Groups II and III and includes pFS2, pFS21, pFS23 and pFS24 which are 6 bp longer than pFS6, the shortest cDNA (Group I). Since half of the sequenced clones were intermediate in length, this length was incorporated into a consensus sequence for the pFS6-like clones, designated pFS6c (Figure 5.1). The remaining clones, pFS8 and pFS19, constitute a class of cDNA, Group IV, which is

23 bp longer than pFS6. The classification of pFS13 is unknown due to an incomplete sequence at the 3' end. Five nucleotide substitutions were observed in pFS24 and a deletion of one nucleotide from the pFS6 sequence was observed (Figure 5.1) compared with the putative consensus sequence.

There are several explanations for the sequence and length divergence between the pFS6 homologues. The length differences may have arisen as a result of cloning artefacts and the single nucleotide differences observed in pFS6 and pFS24 may result from reverse transcriptase errors during cDNA synthesis. In particular, the single nucleotide deletion from pFS6 (Figure 5.1) may have occurred as a result of enzyme slippage during transcription of the gene or reverse transcription of the mRNA. However, since most of the sequence differences occur in pFS24, they are clearly not random, as would be expected if the differences were due to enzyme or sequencing error. Alternatively it may be that the different pFS6-like cDNA classes represent transcripts from distinct but closely-related members of a fibre-specific gene family in Siokra 1-2. The genes may have diverged in both sequence and length, possibly without affecting the coding regions of the genes.

Another explanation is that the three cDNA length variants result from alternative use of the three putative polyadenylation sites within the cDNA sequences. Analysis of the 3' untranslated regions of the cDNAs revealed the presence of a polyadenylation signal sequence 5'-AAUAAA-3', 89 nt upstream from the site of polyadenylation. In addition, there are two divergent polyadenylation signals, 5'-AAUAUU-3' and 5'-AAUUAA-3', which occur 23 nt and 4 nt upstream from the site of polyadenylation respectively (Figure 5.1). Multiple polyadenylation sites are a common feature of plant genes, and in some cases, these are used differentially during development. In genes with multiple poly(A) sites, each site is controlled by a different AAUAAA-like element (Mogen *et al.*, 1992), and it is thought that the relative effectiveness of each motif may be influenced by *cis*-acting sequence elements, located upstream or downstream of the cleavage site. The defining feature of most cleavage and

poly(A) sites in plant genes is probably the occurrence of a YA dinucleotide (Y = pyrimidine) in a U-rich sequence context (reviewed in Hunt, 1994). Such a feature is present in all three putative polyadenylation sites of the consensus pFS6 cDNA sequence.

In contrast to mammalian systems, polyadenylation signals in plant genes are ill-defined and results from several experiments suggest that variation from the canonical signal, AAUAAA, can be tolerated (Wu *et al.*, 1993; Rothnie *et al.*, 1994; Li and Hunt, 1995). In some cases, variant signals are used preferentially (Heidecker and Messing, 1986). Compilational analyses of plant nuclear gene sequences suggest that conservation is restricted to the first three nucleotides only, AAU (Joshi, 1987a). In addition, AAUAAA-like elements in plant genes can be situated very near to their associated polyadenylation sites (Rothnie *et al.*, 1994; Li and Hunt, 1995). Taken together, the data suggests that all three putative polyadenylation signals identified in the pFS6 consensus sequence may be functional. The third putative signal, AAUUAA, may be utilised by the longest (Group IV) transcripts in this sample, namely pFS8 and pFS19 (Figure 5.1).

The pFS6 consensus sequence was used to identify the protein encoded by the group of fibre-specific cDNA clones. The sequence of pFS6c showed significant similarity at the nucleic acid level (GenBank and EMBL; March, 1995) to cDNAs encoding phospholipid transfer proteins (LTPs), a class of plant proteins thought to be involved in the biogenesis of cellular membranes (Arondel and Kader, 1990). In particular, the pFS6c nucleotide sequence is very similar (97%) to that of GH3, a cotton fibre-specific LTP cDNA recently isolated from *Gossypium hirsutum* L., cv. DES119 (Ma *et al.*, 1995; Section 1.6.2). The 58 residue conceptual translation of pFS6c shows 91% amino acid identity to the corresponding region of the GH3 protein, containing five amino acid substitutions of which only three are to amino acids with similar physical properties.

Notably, the cotton fibre-specific LTP cDNA isolated from cv. DES119, GH3 (Ma *et al.*, 1995), can be classified as a member of the Group III pFS6-homologous clones which

contains only pFS24 (Figure 5.1; Table 5.1). The length of GH3 is intermediate, placing it in either Group II or Group III, but the 131 3'-terminal nucleotides of the GH3 sequence is identical to the sequence of pFS24 and differs from the pFS6 consensus sequence. It may be that GH3 and pFS24 are of the same LTP gene subfamily, whilst at least one other family member is defined by the remaining pFS6 clones and the consensus sequence pFS6c.

5.3 Sequence Analysis of pFS6-like Clones with Multiple Inserts

Three of the pFS6 group clones which were sequenced, pFS13, pFS23 and pFS24, contained multiple inserts as a result of cloning artefacts during cDNA synthesis. The pFS13 clone insert was expected to consist of multiple cDNAs since its size of approximately 800 bp (Figure 4.3a) exceeded the estimated transcript size for pFS6 of 650 nt (Figure 4.4a). As was the case for clone pFS14 (Section 4.5.3a), the non-pFS6 inserts do not affect the fibre-specificity of pFS13 as a whole, since the clone was hybridised only by fibre cDNA in the differential screen (not shown). Sequencing of pFS13 yielded 293 bp (designated pFS13F) and 339 bp (designated pFS13R) from the (-40) and RSP of pBluescript®SK(-) respectively. There was no evidence for multiple inserts such as presence of poly(A) tails, *Eco*RI sites or *Not*I sites with associated anchor sequences in either pFS13F or pFS13R and as expected, the sequences did not overlap. The pFS13R sequence (Figure 5.1) was identical to that of pFS6 except that it was truncated at the 3' end, terminating at nucleotide 82 of the pFS6 sequence. Further sequencing would reveal into which of the classes of pFS6-homologous transcripts it falls. The remainder of the pFS13R sequence was identical to that of pFS19 and extended the 5' end of pFS19 by 84 bp (Figure 5.1).

The 293 bp pFS13F sequence (not shown) did not show any sequence similarity to the pFS6 group clones but a search of the nucleotide databases (GenBank and EMBL; January, 1996) revealed significant similarity to ADP-ribosylation factors (ARFs) from a variety of plants. The cotton fibre ARF is homologous to several ARFs isolated from *Arabidopsis*

(Regad *et al.*, 1993; Newman *et al.*, 1994) and rice (Higo *et al.*, 1994; Inoue *et al.*, 1994; Sasaki *et al.*, 1994), in addition to those from *Chlamydomonas reinhardtii* (Memon *et al.*, 1995), potato (Szopa and Müeller-Röeber, 1994), maize (Verwoert *et al.*, 1995) and carrot (Kiyosue and Shinozaki, 1995). The best match was a nucleotide identity of 86% with an *Arabidopsis* ARF mRNA (Regad *et al.*, 1993) which incorporated the first 121 bp of the pFS13F sequence.

ADP-ribosylation factors constitute a family of GTP-binding proteins which have been implicated in vesicle-mediated protein transport between the endoplasmic reticulum and the Golgi compartments of yeast and mammalian cells (Balch *et al.*, 1992). ADP-ribosylation factors are distinct from the *ras* superfamily of GTP-binding proteins which contains two cotton fibre-specific members (Delmer *et al.*, 1995). It has also been postulated that ARFs function as molecular switches that can be turned on by GTP and off by hydrolysis of GTP to GDP. In this way they could control cell processes such as cell proliferation, differentiation and intercellular vesicular transport (Verma *et al.*, 1994). In cells of higher plants, as in other eukaryotic cells, the Golgi apparatus is involved in the transport of storage proteins to the vacuoles. Another function of the Golgi apparatus, unique to plant cells, is the synthesis of complex cell wall polysaccharides, hemicelluloses and pectins (Moore *et al.*, 1991). Plant ARFs and other small GTP-binding proteins should play a key role in these processes and may have one or more of the above functions in cotton fibre cells.

Sequencing of clone pFS23, estimated in size at approximately 700 bp (Figure 4.3a), revealed that it too contained multiple cDNA inserts. The 178 bp derived from the RSP of pBluescript®SK(-), pFS23R, was identical to pFS2, placing it in clone Group II, and included a poly(A) tail of seven adenine residues (Figure 5.1). Extensive cross-banding in the remainder of the pFS23R sequence and in the pFS23 sequence obtained from the (-40) primer of pBluescript®SK(-), presumably derived from an unrelated cDNA or cDNAs, prevented determination of its nature.

Similarly, the 464 bp pFS24 clone contained multiple inserts but without obvious breakpoints between the independent cDNAs. The terminal 131 bp of pFS24 was identical to the sequence of pFS2 except for five nucleotide differences which made the clone unique and placed it in the third group of pFS6 homologues (Figure 5.1). The 261 bp of non-pFS6 sequence (not shown) showed a significant degree of similarity to an anonymous partial cDNA sequence from the *Arabidopsis* expressed sequence tag (EST) collection (Newman *et al.*, 1994) and has no known function. The sequences showed 75% nucleotide identity over a 193 bp overlap.

5.4 Isolation of the Full-length pFS6 cDNA by 5'-RACE

Comparison of cDNA lengths (Figure 4.3a) with an estimated transcript size of 650 nt (Figure 4.4a) indicated that a full-length cDNA was not present in the pFS6-related cDNA clone sample. The remaining portion of the pFS6 mRNA was isolated *via* rapid amplification of cDNA ends (5'-RACE), using oligonucleotide primers complementary to the pFS6 consensus cDNA sequence. The resultant 500 bp 5'-RACE PCR product (Figure 5.2a, b) was cloned and two positive clones, pFS6-30 and pFS6-56 (Figure 5.2c, d), were sequenced in both directions.

An alignment of the 5'-RACE clone sequences (Figure 5.3) includes an additional insert of 111 bp which was excised by *Eco*RI from clone pFS6-56 (Figure 5.2d, lane 3) and is presumably a truncated 5'-RACE product which was cloned during the ligation step of 5'-RACE PCR product cloning. The sequence of pFS6-56b showed a high degree of sequence divergence (five single nucleotide substitutions; Figure 5.3) from the sequence of pFS6-30 and the larger of the two inserts of pFS6-56. Comparison of the pFS6-56b sequence with the pFS6-like cDNA clones (Figure 5.1) revealed identity to pFS24 and GH3 (Ma *et al.*, 1995) in the corresponding region, except for a single nucleotide substitution at position 46 in the pFS6-56b sequence (Figure 5.3). Clearly, insert pFS6-56b is derived from extension of an LTP

mRNA which is a member of the Group III pFS6 homologues, also containing pFS24 and GH3. The minor sequence difference may be attributable to a difference in LTP alleles between the cultivars *G. hirsutum* L., cv. DES119 and *G. hirsutum* L., cv. Siokra 1-2. A 5'-RACE experiment which specifically targeted the pFS24 transcript or the identification of a full-length 5'-RACE product which extends pFS6-56b may result in isolation of a full-length clone identical in sequence to GH3.

Sequencing of the pFS6-30 insert and the larger insert of pFS6-56 both yielded 448 bp of sequence which, as expected, overlapped with the consensus sequence of the pFS6 cDNA (Figure 5.3). In addition, the 5'-RACE derived sequence was identical to that of clone pFS13 (Figure 5.1). The two RACE clones were identical except for a single nucleotide substitution at position 104, which results in substitution of a C in pFS6-30 for a T in pFS6-56 (Figure 5.3). The difference in sequence is most likely due to a *Taq* DNA polymerase error during PCR but it is unknown which of the two clones contains the error. Sequencing of a third 5'-RACE PCR product would resolve the discrepancy.

The full length cDNA, designated cFSltp, was constructed from the sequences of clones pFS6 and pFS6-30 (Figure 5.4). The sequence contains a single, long open reading frame which potentially encodes a 120 amino acid polypeptide, designated FSLTP, with a calculated molecular mass of 11613 Da and a pI of 8.4. The putative initiation codon, at nucleotide position 78 in the cFSltp sequence, contains a purine at the -3 position and a GCT triplet at position +1 relative to the start site, which is a highly conserved feature of plant initiation codons (Heidecker and Messing, 1986; Joshi, 1987b; Lütcke et al, 1987). Two other candidates for initiation codons are in suboptimum context and are both downstream of the putative functional AUG, located at nucleotide positions 90 and 106 in the cFSltp sequence (Figure 5.4). As already discussed, initiation of translation occurs at the first AUG triplet of over 97% of eukaryote mRNAs (Kozak, 1984) and over 90% of plant mRNAs (Heidecker and Messing, 1986; Joshi, 1987b).

Figure 5.2: Cloning of the 5' end of cDNA clone pFS6

a. Agarose gel showing products resulting from PCR amplification of 5'-RACE products of pFS6. Lanes 2-5 are each loaded with a 10 µl sample of PCR product, from a reaction volume of 50 µl and the lanes are:

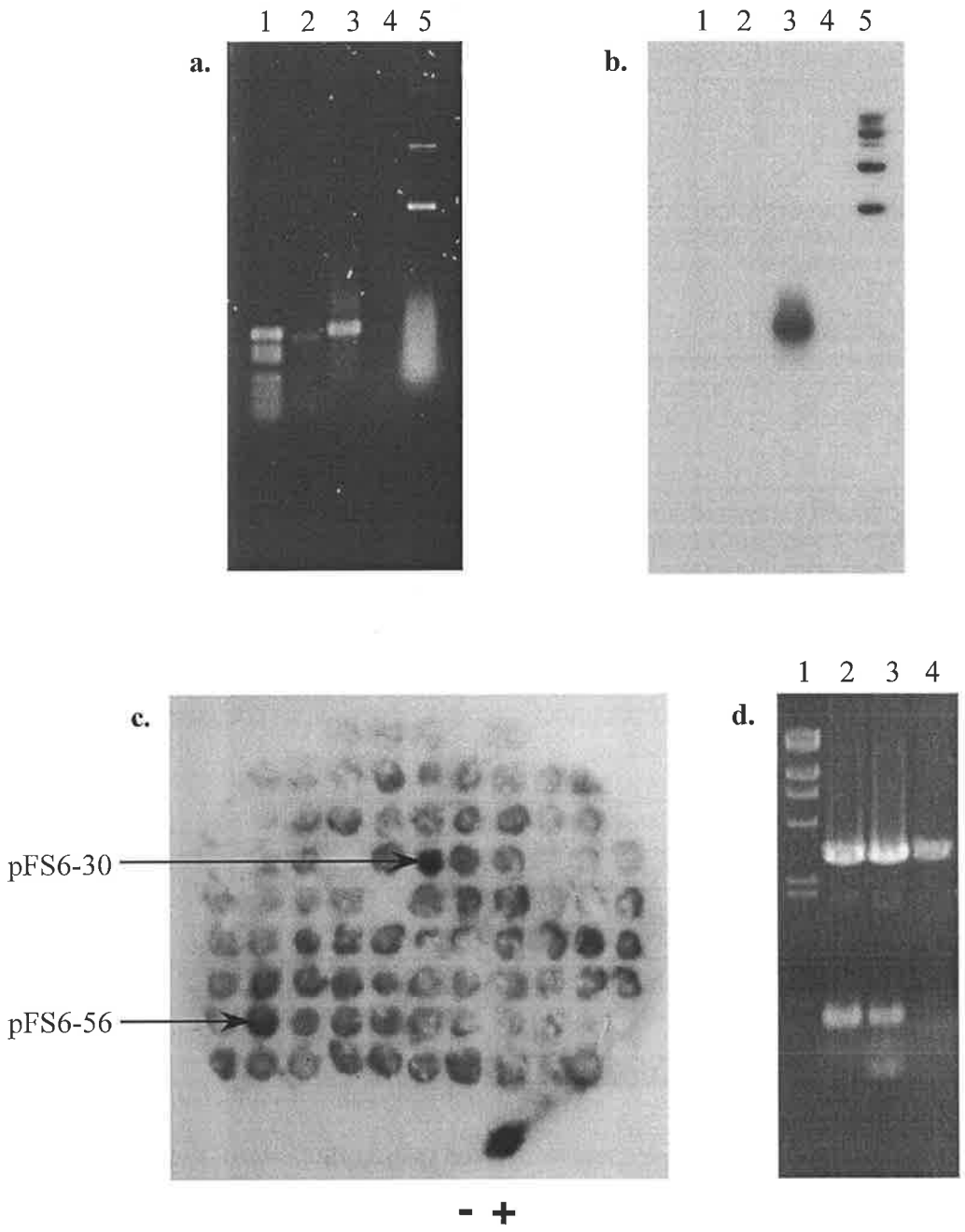
1. Molecular weight markers (pUC19 DNA restricted with *Hpa*II) ,with fragment sizes 501 bp, 489 bp, 404 bp 331 bp, 242 bp, 190 bp, 147 bp and 111 bp;
2. TFR control cDNA after both purification and anchor ligation, utilising one gene-specific primer together with the anchor primer. A product of 480 bp should be amplified;
3. 5'-RACE PCR of pFS6 cDNA, utilising gene-specific PCR primer P21.2 together with the anchor primer;
4. Negative control reaction, using the same primers as those used in the experiment, but without the addition of a cDNA template;
5. A 1 µg sample of uncut pFS2 plasmid DNA, included as a positive control for the subsequent Southern analysis.

b. Southern blot of the gel in part a., hybridised with the pFS2 cDNA insert. Lanes are numbered as in part a.

c. Colony hybridisation of putative pFS6 5'-RACE clones, probed with the pFS2 insert. Positive clones pFS6-30 and pFS6-56 are indicated by arrows, and the positions of positive and negative control colonies are indicated by a (+) and (-) respectively.

d. Agarose gel showing the pFS6 5'-RACE clones identified above. Each lane is loaded with approximately 2 µg of plasmid DNA restricted with *Eco*RI and the lanes are:

1. Molecular weight markers (Lambda DNA restricted with *Hin*dIII), with fragment sizes 23.1 kb, 9.4 kb, 6.6 kb, 4.4 kb, 2.3 kb and 2.0 kb;
2. Clone pFS6-30;
3. Clone pFS6-56;
4. A 1 µg sample of pBluescript®SK(-), restricted with *Eco*RI.


1 2 3 4 5
a.
1 2 3 4 5
b.
c.
pFS6-30
pFS6-56
- +
1 2 3 4
d.

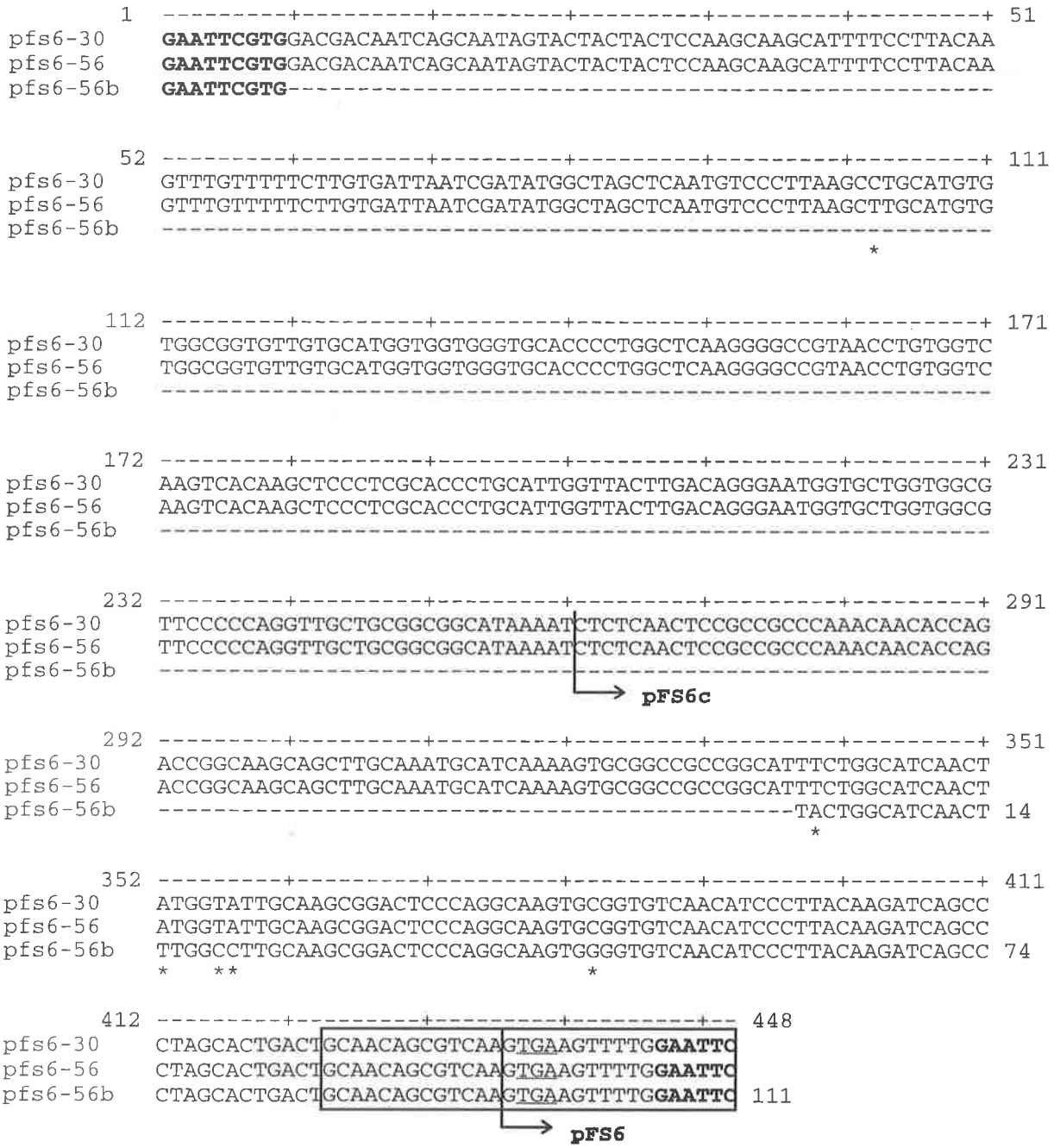


Figure 5.3: Alignment of pFS6 5'-RACE clone sequences

Alignment of the three insert sequences derived from pFS6 5'-RACE clones pFS6-30 and pFS6-56 ~~together with a consensus sequence~~. The region complementary to the 5'-RACE PCR primer sequence P21.2 is boxed and *Eco*RI restriction sites 5'-GAATTC-3' are in bold, as are the anchor sequences 5'GTG-3'. Putative TGA stop codons in each sequence are underlined. The 5' end of the pFS6c and pFS6 sequences are indicated by arrows, and sequence differences between the 5'-RACE clones are indicated by asterisks below the sequence alignment. Numbers refer to nucleotide positions in the insert sequences.

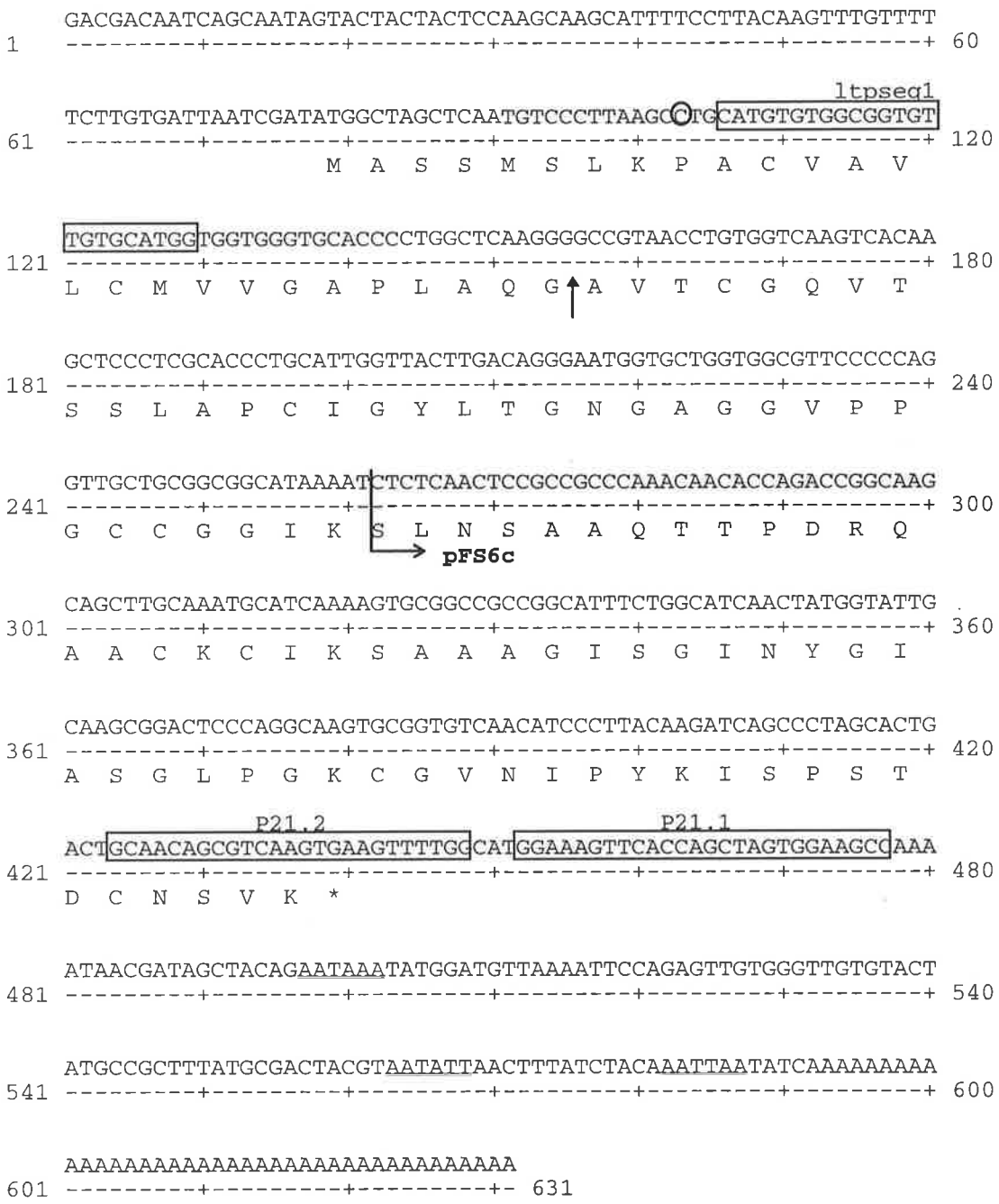


Figure 5.4: Sequence of the full-length pFS6 fibre-specific cDNA

Nucleotide sequence and deduced amino acid sequence of the full-length cotton LTP cDNA, cFSltp, constructed from the pFS6 cDNA consensus sequence and the sequence of the pFS6-30 clone. The putative signal peptide cleavage site is marked with an arrow and the pFS6c sequence indicated. The regions complementary to the oligonucleotide primers used for 5'-RACE and primer ltpseq1 are boxed and putative polyadenylation signal(s) are double underlined. The nucleotide which differed between the two 5'-RACE clone sequences is circled.

The nucleotide sequence of the full-length cDNA clone cFSltp shows 95% identity to that of GH3, a cotton fibre-specific LTP cDNA previously isolated from *G. hirsutum* L., cv. DES119 (Ma *et al.*, 1995; Figure 5.5). The 120 amino acid conceptual translation of cFSltp shows 88% amino acid identity to GH3, containing 14 amino acid substitutions, 10 of which are conservative changes. One of the differences corresponds to the single bp mismatch between the 5'-RACE clones. The GH3 sequence matches that of pFS6-56, having a T at nucleotide position 104, instead of the C present in RACE clone pFS6-30 (Figure 5.5). It may be that the remaining 31 single nucleotide changes are attributable to allelic differences between the cultivars *G. hirsutum* L., cv. DES119 and *G. hirsutum* L., cv. Siokra 1-2, but results from this study suggest that fibre-specific LTPs form a small subfamily, of which cFSltp and GH3 are both members. One of the fibre-specific cDNA clones, pFS24, may be the GH3 homologue in Siokra 1-2 and it appears that a partial 5'-RACE product, pFS6-56b, is an extension of pFS24 (Figure 5.5). The cFSltp class of transcript is defined by the majority of the pFS6-homologues together with the 5'-RACE clones pFS6-30 and pFS6-56, while GH3, pFS24 and pFS6-56b define a second class of fibre-specific LTP. Different LTP isoforms have been isolated from broccoli (Pyee and Kolattukudy, 1995), rape seedlings (Tsuboi *et al.*, 1991) and tobacco (Koltunow *et al.*, 1990; Fleming *et al.*, 1992), and these isoforms also differ in their expression patterns. However, two LTP genes isolated from barley (Skriver *et al.*, 1992; Kalla *et al.*, 1994) differ in sequence but both exhibit aleurone-specific expression and they may belong to the same LTP gene family.

In addition, the cotton fibre LTP sequence is similar to those found in spinach (Bernhard *et al.*, 1991), tobacco (Fleming *et al.*, 1992), tomato (Torres-Schumann *et al.*, 1992), broccoli (Pyee *et al.*, 1994) and carrot (Sterk *et al.*, 1991) with which it shares amino acid identities of 77%, 64%, 63%, 57% and 54% respectively (Figure 5.6a). In particular, the eight cysteine residues are strictly conserved (Figure 5.6a), which makes it likely that the proteins fold into analogous three-dimensional structures, mediated by disulphide bridges (Tchang *et al.*, 1988;

Shin *et al.*, 1995). In addition to the conserved cysteines, some sequence stretches and relative positions of cysteine and proline residues are highly conserved among LTPs and may play an important role in their structure and function. The charged portion ^{72}DRQ^{74} and the hydrophobic amino acid sequence ^{98}LPGKCGVNIPY^{108} (numbers refer to FSLTP) have been described as probable sites of interaction for the phospholipid phosphate group and the acyl chains respectively (Tchang *et al.*, 1988). The conceptual amino acid sequence of cFSltp also shows a high degree of sequence similarity with LTPs from monocots such as rice (Vignols *et al.*, 1994; 53% identity) and barley (Linnestad *et al.*, 1991; 43% identity). The high conservation among the LTP protein sequences from several dicots and monocots suggests that the LTP gene family arose prior to the evolutionary divergence of monocots and dicots some 150 million years ago.

After the AUG translation initiation codon, there is a stretch of 26 amino acids which contains many of the features characteristic of signal peptides (Nothwehr and Gordon, 1990). The putative cleavage site of the signal peptide was predicted according to von Heijne (1983) and confirmed by alignment with homologues from other plants (Figure 5.6a). Assuming that cleavage occurs between amino acids 26 and 27 of FSLTP (Figure 5.4; Figure 5.6a), the mature protein consists of 94 residues with a calculated molecular weight of 9092 Da and a basic pI of 8.4, which is consistent with data reported for other LTPs.

A hydropathy plot of FSLTP revealed a generally hydrophobic protein, with only 2-3 hydrophilic regions and few charged residues, comparable to hydropathy plots of other plant LTPs (Figure 5.6b). Also apparent is the N-terminal hydrophobic domain which is typically associated with signal peptides and present in all LTPs characterised to date. This pattern is very consistent in the tobacco homologue, with which the cFSltp open reading frame shares 64% amino acid sequence similarity, and also in carrot to which it is less similar (54%). The other cotton LTP (GH3; Ma *et al.*, 1995) is by far the most similar to the mature FSLTP protein at the amino acid level (87%) and yet appears significantly different from FSLTP and

Figure 5.5: Alignment of fibre-specific cDNAs for lipid transfer proteins

Nucleotide sequence comparison between two fibre-specific cDNAs which encode LTPs, together with the sequence of cDNA clone pFS24 and 5'-RACE product pFS6-56b. Differences are indicated by asterisks below the sequence alignment. GH3 was isolated from *G. hirsutum* L., cv. DES119 (Ma *et al.*, 1995) and the remaining sequences were derived from clones isolated from *G. hirsutum* L., cv. Siokra 1-2 in this study. Numbering of nucleotides on the scale refers to cFSltp and GH3. The sequence of pFS24 is identical to GH3 in the region of overlap, as is pFS6-56b with the exception of one nucleotide. The sequences of GH3 and cFSltp diverge at 32 nucleotide positions.

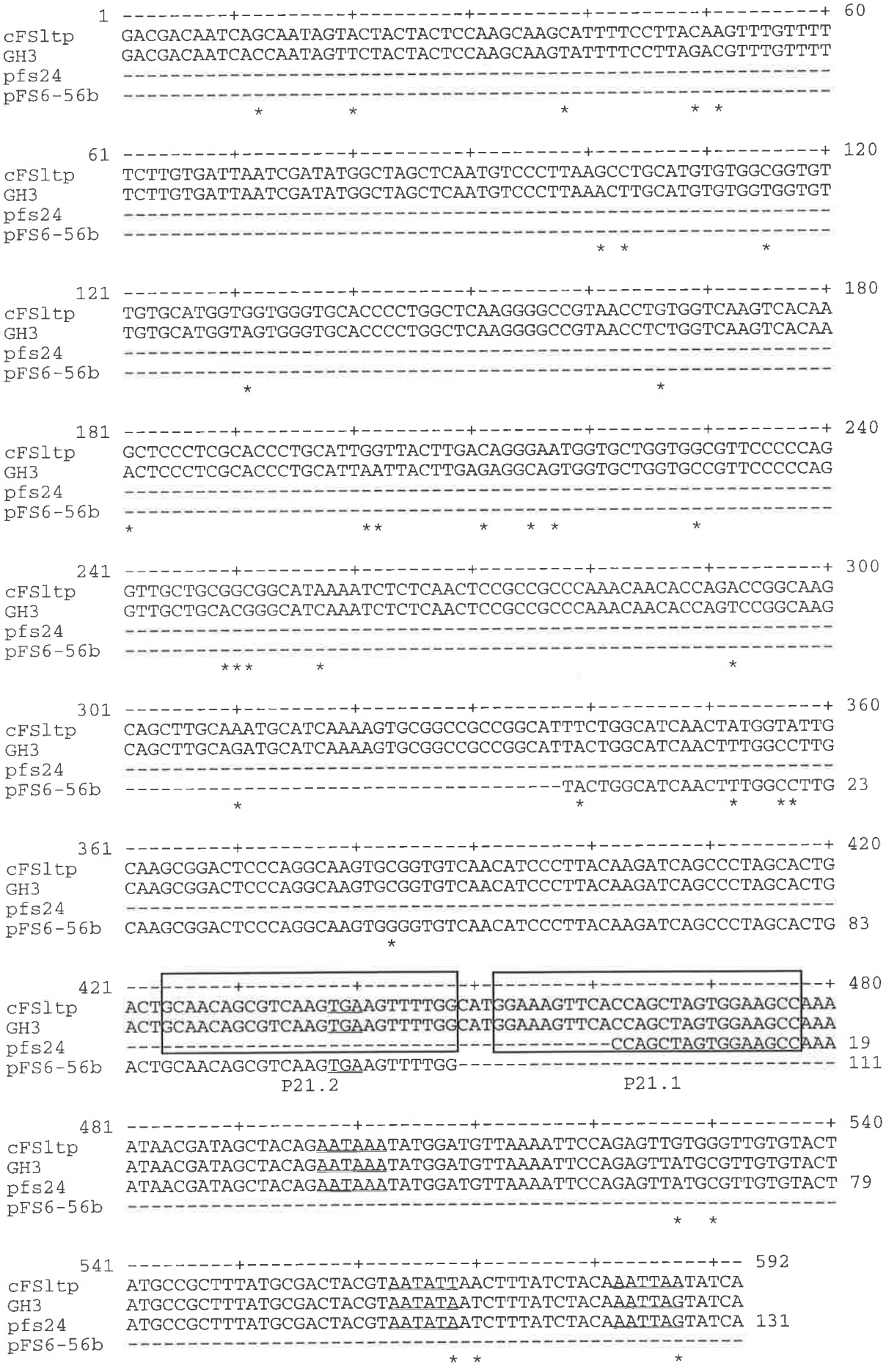
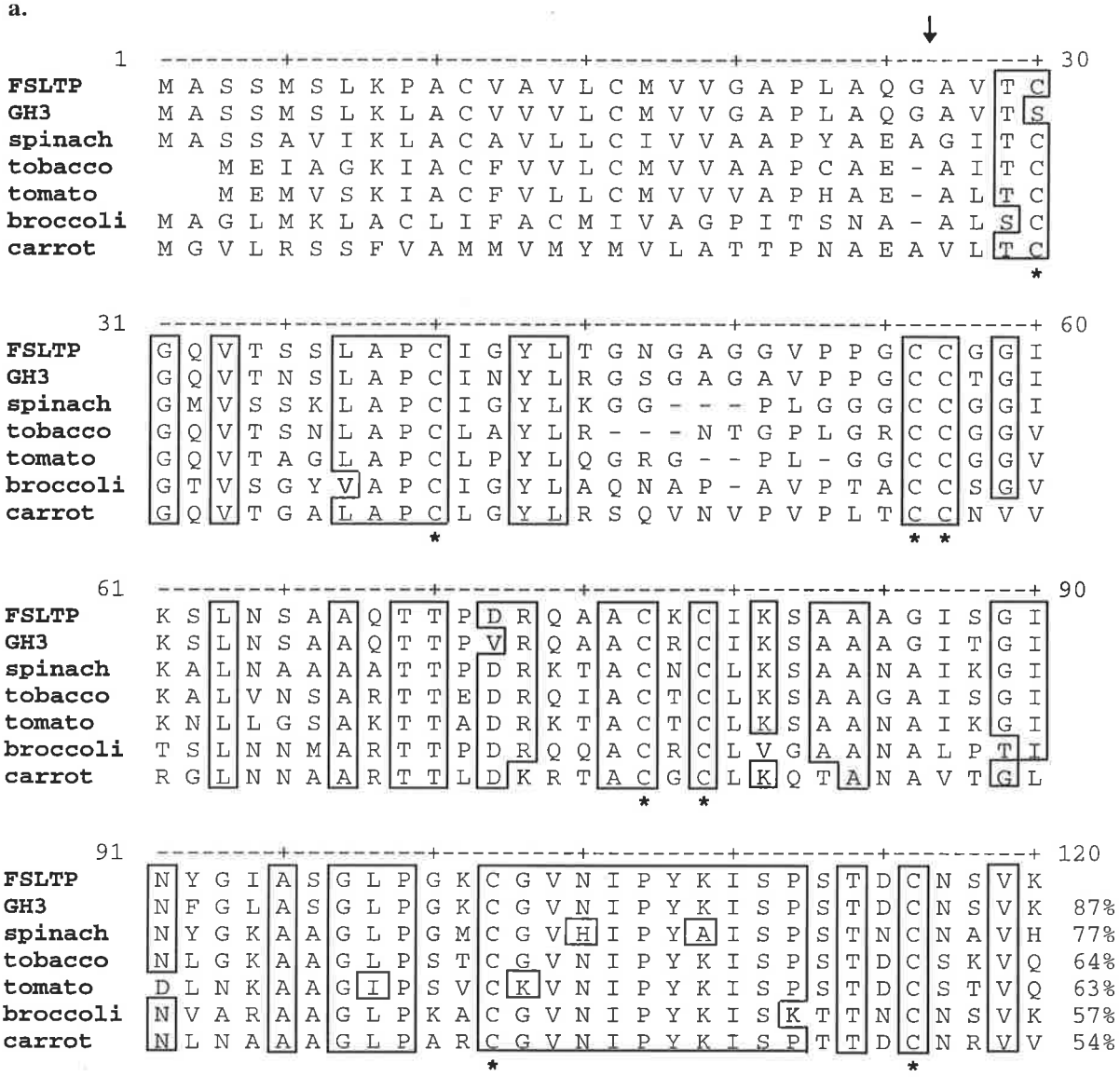

cFSltp
GH3
pfs24
pFS6-56b
P21.2
P21.1

Figure 5.6: Analysis of the 120 amino acid FSLTP protein

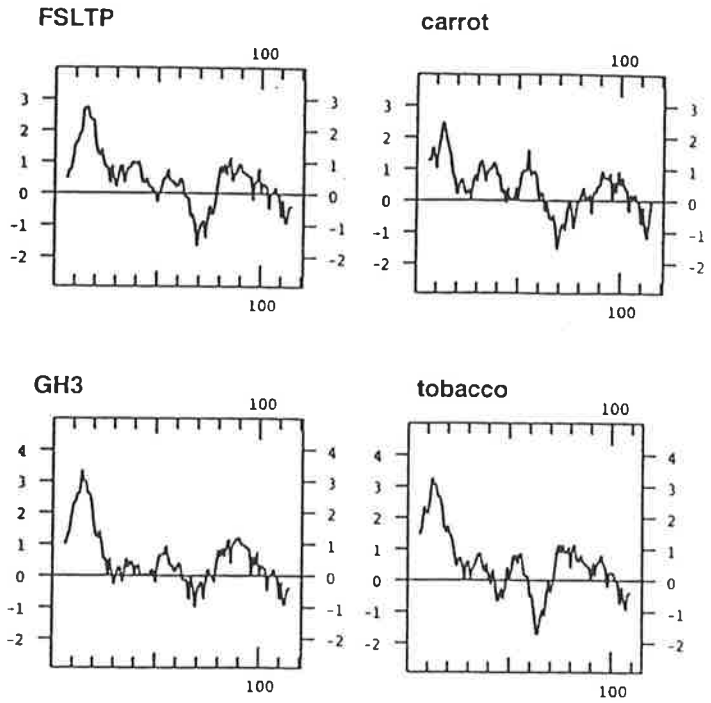
a. Sequence comparison of seven plant lipid transfer proteins. Boxes indicate homologous amino acids in at least six of the sequences and homology with *G. hirsutum* L. FSLTP (mature peptide only) is indicated at the end of each sequence. Putative signal peptide cleavage sites are marked with an arrow and the highly conserved cysteine residues are indicated with asterisks.

b. Comparison of the hydropathy plots of two cotton fibre-specific LTPs with the homologues from carrot and tobacco. The horizontal scale indicates the number of amino acid residues and the vertical one the relative scale, according to Kyte and Doolittle (1982). Points above the horizontal line correspond to hydrophobic region and points below this line are hydrophilic.

a.


1 30
FSLTP M A S S M S L K P A C V A V L C M V V G A P L A Q G A V T C
GH3 M A S S M S L K L A C V V V L C M V V G A P L A Q G A V T S
spinach M A S S A V I K L A C A V L L C I V V A A P Y A E A G I T C
tobacco M E I A G K I A C F V V L C M V V A A P C A E - A I T C
tomato M E M V S K I A C F V L L C M V V V A P H A E - A L T C
broccoli M A G L M K L A C L I F A C M I V A G P I T S N A - A L S C
carrot M G V L R S S F V A M M V M Y M V L A T T P N A E A V L T C
31 60
FSLTP G Q V T S S L A P C I G Y L T G N G A G G V P P G C C G G I
GH3 G Q V T N S L A P C I N Y L R G S G A G A V P P G C C T G I
spinach G M V S S K L A P C I G Y L K G G - - - P L G G G C C G G I
tobacco G Q V T S N L A P C L A Y L R - - - N T G P L G R C C G G V
tomato G Q V T A G L A P C L P Y L Q G R G - - P L - G G C C G G V
broccoli G T V S G Y V A P C I G Y L A Q N A P - A V P T A C C S G V
carrot G Q V T G A L A P C L G Y L R S Q V N V P V P L T C C N V V
61 90
FSLTP K S L N S A A Q T T P D R Q A A C K C I K S A A A G I S G I
GH3 K S L N S A A Q T T P V R Q A A C R C I K S A A A G I T G I
spinach K A L N A A A A T T P D R K T A C N C L K S A A N A I K G I
tobacco K A L V N S A R T T E D R Q I A C T C L K S A A G A I S G I
tomato K N L L G S A K T T A D R K T A C T C L K S A A N A I K G I
broccoli T S L N N M A R T T P D R Q Q A C R C L V G A A N A L P T I
carrot R G L N N A A R T T L D K R T A C G C L K Q T A N A V T G L
91 120
FSLTP N Y G I A S G L P G K C G V N I P Y K I S P S T D C N S V K
GH3 N F G L A S G L P G K C G V N I P Y K I S P S T D C N S V K 87%
spinach N Y G K A A G L P G M C G V H I P Y A I S P S T N C N A V H 77%
tobacco N L G K A A G L P S T C G V N I P Y K I S P S T D C S K V Q 64%
tomato D L N K A A G I P S V C K V N I P Y K I S P S T D C S T V Q 63%
broccoli N V A R A A G L P K A C G V N I P Y K I S K T T N C N S V K 57%
carrot N L N A A A G L P A R C G V N I P Y K I S P T T D C N R V V 54%

b.


FSLTP
carrot
GH3
tobacco
100

the other LTPs in hydropathy profile. It may be that sequencing errors contribute to changes in the amino acid sequence of the GH3 protein, producing an atypical hydropathy profile for a generalised LTP.

5.5 Lipid Transfer Proteins in Plants

Lipid transfer proteins have been characterised in a number of higher plants and have several features in common which are shared by the cotton fibre LTPs. These include low molecular mass (9 kDa), high (basic) isoelectric point and high sequence conservation. Plant LTPs show a broad specificity for lipid substrates and are abundant proteins which may comprise up to 4% of the total soluble protein within a cell (Kader, 1990). Northern analysis of plant LTP genes reveals complex expression patterns which are often characterised by a high tissue specificity. The fibre-abundant expression pattern of two LTP genes in cotton is another example which demonstrates the diversity of expression patterns for lipid transfer proteins in plants. For instance, there are LTPs whose expression is restricted to the tapetal cells of the anther (Koltunow *et al.*, 1990), to epidermal cells of somatic embryos (Sterk *et al.*, 1991), to petal and carpel tissues (Kotilainen *et al.*, 1994), or which are induced during seed germination (Linnestad *et al.*, 1991; Torres-Schumann *et al.*, 1992). Plant LTPs are generally encoded by multigene families of at least two non-allelic members, with different genes of the family encoding similar proteins that may have distinct expression patterns and/or functions.

Lipid transfer proteins are characterised by their ability to catalyse the exchange of lipids between natural or artificial membranes *in vitro* (reviewed in Kader, 1990). Because of their ability to transfer lipids *in vitro*, LTPs have been assumed to play a role in membrane biogenesis by mediating transport of phospholipids from their site of synthesis to organellar and cellular membranes. However, a cytoplasmic role for LTPs is not easily reconciled with presence of the signal peptide. A number of different functions have been attributed to LTPs, including stress response associated with low temperatures (Hughes *et al.*, 1992), drought

(Plant *et al.*, 1991) or salt treatments (Torres-Schumann *et al.*, 1992). LTPs may also have a defense activity as potent inhibitors of bacterial and fungal plant pathogens (Molina and García-Olmedo, 1993; Segura *et al.*, 1993; Cammue *et al.*, 1995). It is probable that different LTP isoforms fulfill different functions which are not mutually exclusive and which relate to the expression pattern of the particular LTP gene in plant cells. Members of an LTP family in wheat are phosphorylated by a plant calcium-dependant protein kinase which may relate to protein processing and function, lipid transfer function or the anti-microbial function of the protein (Neumann *et al.*, 1994).

Secretion of the protein has been demonstrated in the cases of spinach (Bernhard *et al.*, 1991), barley (Mundy and Rogers, 1986) and carrot (Sterk *et al.*, 1991). LTPs isolated from both *Arabidopsis* and broccoli have been located within the cell wall and in particular, the waxy cuticle (Pyee *et al.*, 1994; Thoma *et al.*, 1993). A role for LTPs in the transport of cutin monomers to the outer epidermal cell wall was therefore suggested (Pyee *et al.*, 1994; Sterk *et al.*, 1991; Thoma *et al.*, 1993). Cutin is the insoluble polymeric structural component of the cuticle of most aerial parts of plants, serving to prevent water loss from epidermal cells and providing a mechanical barrier to pathogen ingress (Kolattukudy, 1981). It is composed of esterified hydroxylated fatty acids, and cutin monomers, derived mainly from palmatic and/or oleic acid, are likely to be transferred to the growing cutin polymer in the form of acyl-CoA esters (Kolattukudy, 1981). LTPs isolated from rape seedlings (Østergaard *et al.*, 1993; Østergaard *et al.*, 1995) and carrot embryo cultures (Meijer *et al.*, 1993) have been demonstrated to bind both fatty acids and their acyl-CoA derivatives, supportive of a function for LTPs in cutin synthesis.

The induction of an analogous activity during cotton fibre development may be related to the extensive cutin deposition which occurs during the elongation phase of growth (Seagull, 1993; Section 1.3.3a) and is consistent with an abundance of pFS6 transcripts early in fibre development (Figure 4.4a). The protective cuticle is absent in other cotton cell types except

epidermal cells. LTP gene expression in epidermal cells may account for the low levels of LTP transcripts present in leaf, flower and seedling tissue.

5.6 Isolation of Putative pFS6 Genomic Clones

Screening of a *G. hirsutum* cv. Siokra 1-2 partial genomic library of 1 x 10^5 recombinants with pFS6 yielded two independent positive clones, λFS6(A) and λFS6(B), in addition to λS5. As with cDNA clones pFS17 and pFS18, hybridisation of pFS6 to λS5 is probably spurious and results from the AT-richness of the sequences or perhaps the poly(A) tails (see also Sections 4.5.3g and 4.5.4b). Hybridisation of pFS6 was much stronger to λFS6(A) and λFS6(B) than to λS5, both on a Southern blot of the three restricted genomic clones and in the genomic library screen (not shown). Restriction of DNA samples from both positive clones with *Eco*RI, *Sac*I, *Xba*I, *Xho*I and *Bam*HI showed that all enzymes contained recognition sites within the inserts except for *Bam*HI (not shown). Double and triple digests using combinations of the four restriction enzymes indicated that clones λFS6(A) and λFS6(B) were virtually identical, with minor differences in the restriction fragment sizes. Restriction mapping of both genomic clones confirmed that λFS6(A) and λFS6(B) are independent isolates of the same clone which are oriented in the opposite directions in the λGEM-11™ vector.

Probing of a Southern of λFS6(B) DNA with the pFS6 insert confined the region of homology to the 2.2 kb *Xho*I fragment, the 12 kb *Sac*I fragment and the 4.9 *Eco*RI fragment of λFS6(B), the latter of which is attached to the long arm of λGEM-11™ (Figure 5.7b; Figure 5.8). When a longer pFS6 homologue, pFS2, was used as a probe to the same Southern, it hybridised several restriction fragments of λFS6(B) in addition to those detected by pFS6 (Figure 5.7c). The additional homology is confined to a 2.7 kb *Xho*I/*Eco*RI fragment (Figure 5.8) by virtue of pFS2 hybridisation to the 4.8 kb *Eco*RI fragment, the 12 kb *Sac*I fragment

and the 5.3 kb *Xho*I fragment of λFS6(B) (Figure 5.7c, lanes 1,2 and 4). The presence of two regions in the genomic clone with homology to pFS2 was suggestive of a large intron near the 3' end of a single LTP gene or the presence of two distinct LTP genes separated by a distance of at least 4.8 kb, as determined by restriction mapping (Figure 5.8). LTP genes often contain an intron near the stop codon, but the longest intron thus far observed is 980 bp in *ltp1* of tobacco, which interrupts two exons of 112 amino acids and 2 amino acids (Fleming *et al.*, 1992). Probing of λFS6(A) Southerns with pFS6 and pFS2 produced identical hybridisation patterns to those depicted in Figure 5.7 (not shown).

In addition, hybridisation of λFS6(B) with a probe generated from the extreme 5' end of cFSltp (not shown) resulted in a pattern which was identical to the one produced by probing with pFS2 (Figure 5.7c). In summary, both the 2.2 kb *Xho*I and 2.7 kb *Xho*I/*Eco*RI fragments of λFS6(B) contain sequences similar to the 5' end of an LTP cDNA, but only one of these, X2.2, shows similarity to the 3' end of the cDNA (pFS6). It can be deduced that the 18 kb genomic DNA fragment cloned in λFS6(B) contains at least two LTP genes separated by a distance of at least 4.8 kb. The 2.2 kb *Xho*I fragment contains sequences homologous to pFS6, whereas the 2.7 kb *Xho*I/*Eco*RI fragment also contains an LTP gene but one which is truncated at the 3' end or shows sequence divergence after the putative stop codon. LTP genes contained within genomic fragments X2.2 and XE2.7 were designated *FSltp1* and *FSltp2* respectively. A similar tandem arrangement of LTP genes has been observed in the genomes of both tobacco (Fleming *et al.*, 1992) and *Sorghum* (Pelèse-Siebenbourg *et al.*, 1994).

5.7 Cloning of Subgenomic DNA Fragments

In order to define the structure and organisation of the putative LTP genes, the fragments of λFS6(B) that hybridised to the pFS2 cDNA were subcloned into a plasmid vector (Figure 5.9a). The genomic fragments which lie adjacent to the regions of interest were also subcloned (Figure 5.9b, c and d), since they may contain promoter sequences which

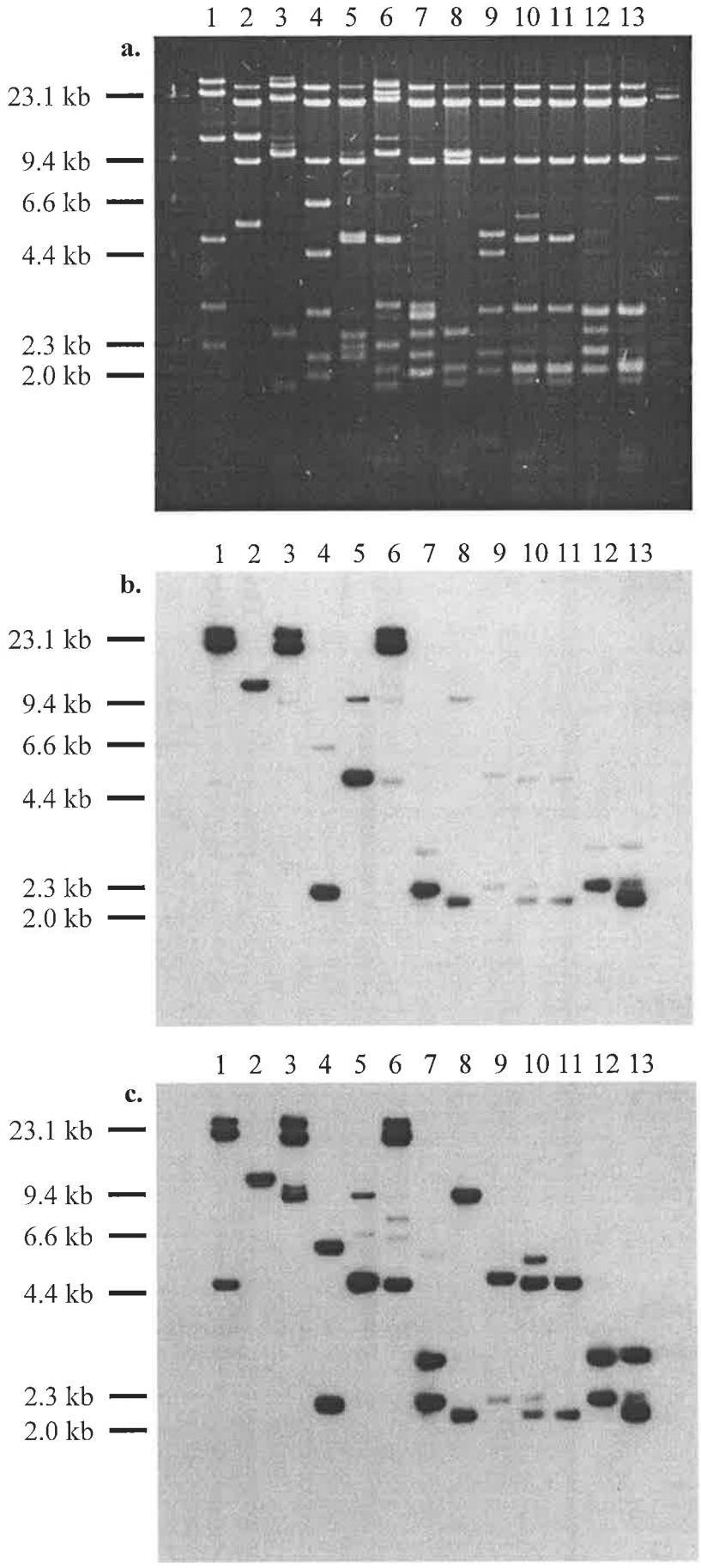
Figure 5.7: Physical mapping of LTP sequences to genomic clone λFS6(B)

a. Agarose gel showing DNA fragments resulting from restriction of λFS6(B). The first and last lanes contain molecular weight markers (Lambda DNA restricted with *Hin*dIII) with sizes as indicated in kb. The remaining lanes each contain approximately 3 μg of DNA digested with the following restriction endonucleases:

1. *Eco*RI;
2. *Sac*I;
3. *Xba*I;
4. *Xho*I;
5. *Eco*RI + *Sac*I;
6. *Eco*RI + *Xba*I;
7. *Eco*RI + *Xho*I;
8. *Sac*I + *Xba*I;
9. *Sac*I + *Xho*I;
10. *Xba*I + *Xho*I;
11. *Sac*I + *Xho*I + *Xba*I;
12. *Eco*RI + *Xho*I + *Sac*I;
13. *Eco*RI + *Xho*I + *Xba*I.

b. Southern blot of gel depicted in a., hybridised with the insert from clone pFS6. Lanes are labelled as above.

c. Southern blot of gel depicted in a., hybridised with the insert from clone pFS2. Lanes are labelled as above.


1 2 3 4 5 6 7 8 9 10 11 12 13
a.
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
1 2 3 4 5 6 7 8 9 10 11 12 13
b.
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
1 2 3 4 5 6 7 8 9 10 11 12 13
c.
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb

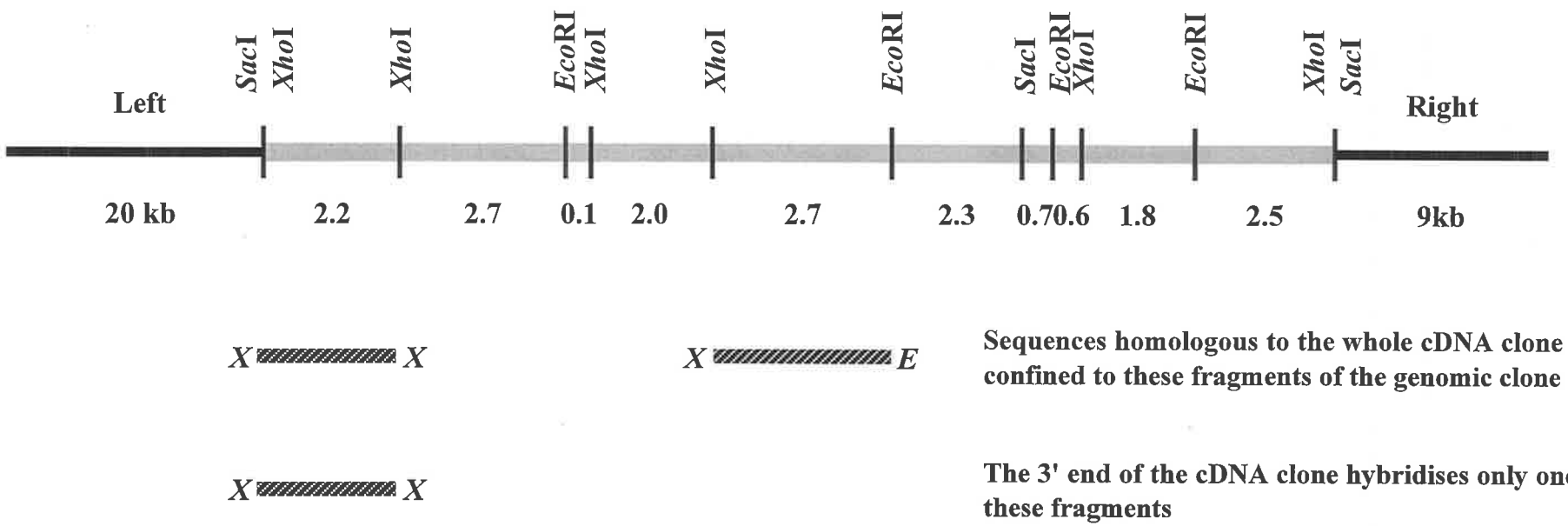


Figure 5.8: Partial restriction map of genomic clone λFS6(B)

Restriction map of genomic clone λFS6(B), showing the positions of the *Xho*I, *Eco*RI and *Sac*I sites. The positions of the XbaI sites were unable to be mapped accurately and are not included in the figure.

correspond to the putative LTP genes. Prior to sequence analysis, the polarity of the LTP genes in each subclone and in the genomic clones was determined using PCR (Figure 5.10). Primers ltpseq1, P21.1 and P21.2 (Table 2.3b) were used in combination with pBluescript®SK(-) primers T3 and T7 (Table 2.3a). PCR using λFS6(B) DNA utilised a T7 priming site in the long arm of λGEM-11™, which is adjacent to the X2.2 fragment of λFS6(B) (Figure 5.8). The results are summarised below.

Table 5.2: Determination of LTP gene orientation in genomic clone and subclones

Clone	Primer combination for PCR	Approximate size of major amplification product (kb)	Sequence amplified
pXE2.7	ltpseq1/T3	1.2	5' coding region and promoter of *FSltp2*
	ltpseq1/T7	-	-
	P21.2/T3	1.5	?
	P21.2/T7	1.5	?
	P21.1/T3	-	-
	P21.1/T7	-	-
	T3/T7	3.0	entire insert of plasmid subclone
pX2.2	ltpseq1/T3	1.1	5' coding region and promoter of *FSltp1*
	ltpseq1/T7	-	-
	P21.2/T3	1.5	?
	P21.2/T7	1.5	?
	P21.1/T3	1.5	coding region and promoter of *FSltp1*
	P21.1/T7	0.7	? non-specific primer binding
	T3/T7	2.2	entire insert of plasmid subclone
λFS6(B)	ltpseq1/T3	1.5	? non-specific primer binding
	ltpseq1/T7	-	-
	P21.2/T3	1.5	?
	P21.2/T7	1.5	?
	P21.1/T3	-	-
	P21.1/T7	-	-
	T3/T7	-	-

Priming of PCR from plasmid subclones pX2.2 and pXE2.7 with T3 in combination with

Figure 5.9: Subcloning of λFS6(B) restriction fragments

The colony hybridisations shown illustrate several strategies for subcloning of λFS6(B) restriction fragments. The top row of colonies corresponds to the gel-purified 2.2 kb *Xho*I fragment of λFS6(B), cloned into pBluescript®SK(-) and the second row consists of colonies resulting from λFS6(B) DNA restricted with *Xho*I and cloned into pBluescript®SK(-). The remaining clones, with the exception of negative and positive controls which are indicated by a (-) and (+) respectively, result from restriction of λFS6(B) DNA with both *Eco*RI and *Xho*I followed by cloning into pBluescript®SK(-).

a. Colony hybridisation probed with the insert from clone pFS6 together with the 5'-RACE clone pFS6-30 insert in order to identify subclones which contain LTP sequences. Several positives are apparent but the subclone designated pX2.2 corresponds to clone #9 and the subclone designated pXE2.7 is clone #28.

b. Colony hybridisation from part a. probed with the 2.7 kb *Xho*I doublet fragment of λFS6(B) in order to identify the 2.8 kb *Xho*I or 2.7 kb *Xho*I/*Eco*RI fragment which flanks pX2.2. Several positive clones were identified (arrowed), based on their hybridisation to this probe but absence of hybridisation to the full-length cDNA probe (above). Subclone pEX2.7 corresponds to clone #88.

c. Colony hybridisation probed with the *Xho*I 2.0 kb fragment of λFS6(B) in order to identify sequences which flank pXE2.7. One positive clone, designated pX2.0 (clone #17) is indicated by an arrow.

d. Colony hybridisation using the 6.3 kb *Xho*I fragment of λFS6(B) in order to subclone the sequences adjacent to pXE2.7 and on the opposite side to pX2.0. Several clones which hybridised to this probe in addition to the full-length cDNA probe were identified (arrowed) and the subclone, designated pEX3.6, corresponds to Grunsteins clone #163.

e. Schematic diagram of λFS6(B), showing partial restriction map and relative positions of subcloned DNA fragments.

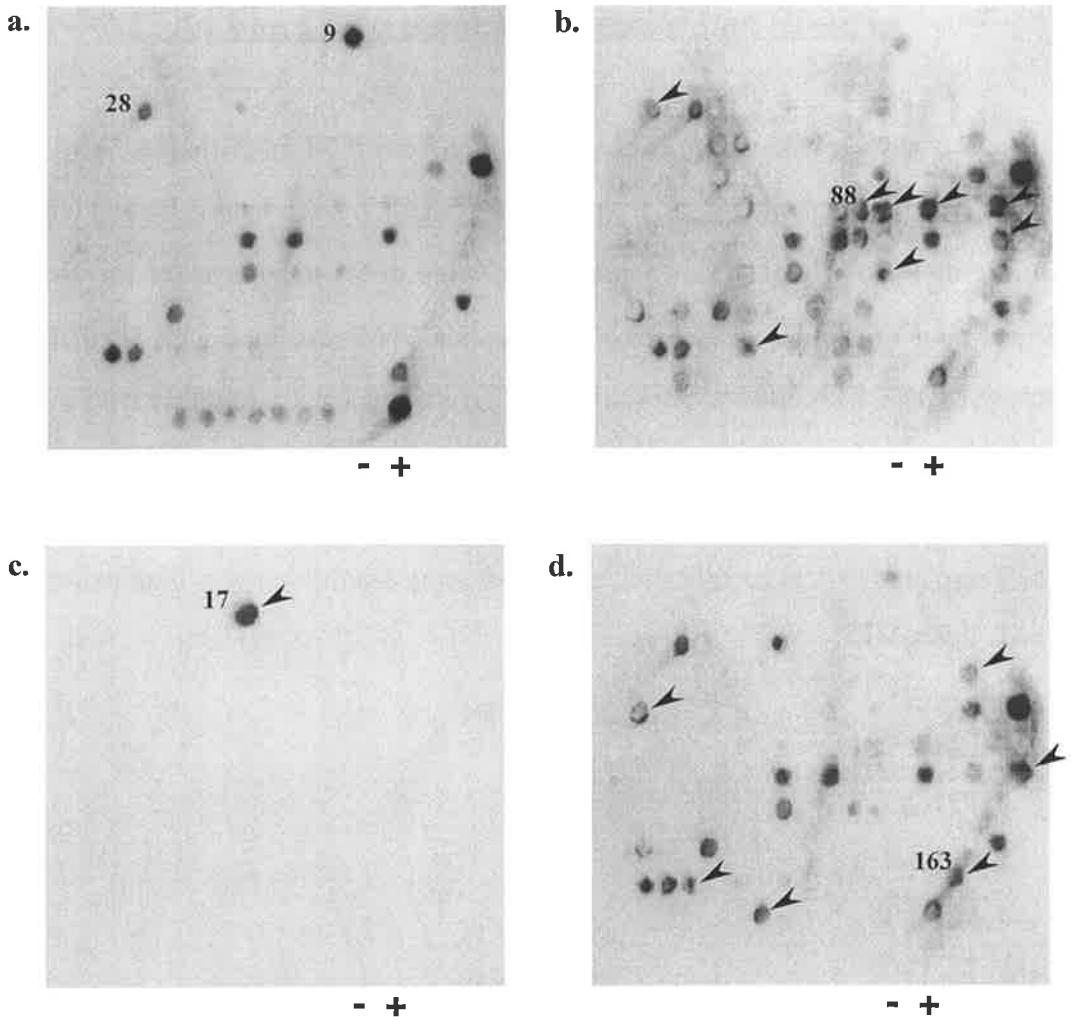

a.
9
28
- +
b.
88
- +
c.
17
- +
d.
163
- +
pX2.2
X X
pXE2.7
X E
e.
SacI XhoI XhoI EcoRI XhoI XhoI EcoRI SacI EcoRI XhoI EcoRI XhoI SacI
2.2 2.7 0.1 2.0 2.7 2.3 0.7 0.6 1.8 2.5
X E
pEX2.7
E X
pEX3.6
X X
pX2.0

Figure 5.10: Determination of LTP gene orientation in genomic clone λFS6(B) and plasmid subclones pX2.2 and pXE2.7

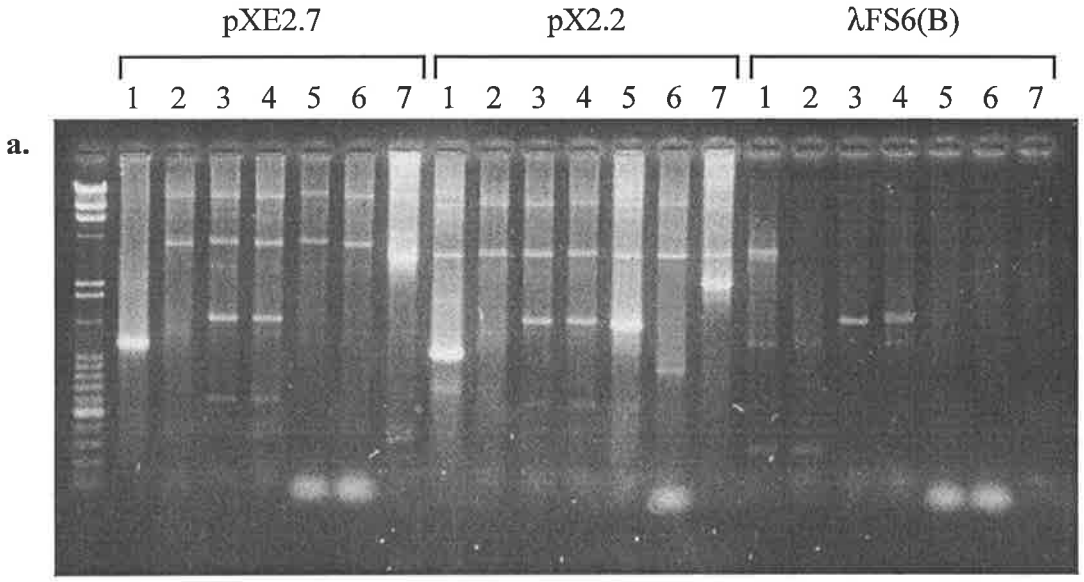
a. Agarose gel showing products resulting from PCR amplification of genomic clone λFS6(B) and plasmid subclones pX2.2 and pXE2.7. Each lane is loaded with a 10 µl sample of PCR product from a reaction volume of 25 µl except for the first lane which contains molecular weight markers (Lambda DNA restricted with *Hin*dIII), with fragment sizes 23.1 kb, 9.4 kb, 6.6 kb, 4.4 kb, 2.3 kb and 2.0 kb, together with a 100 bp ladder which consists of fragments ranging in size from 100 bp to 1000 bp in 100 bp increments, plus an additional fragment of 1500 bp.

Template DNA is as indicated, amplified using the following primer pairs:

1. ltpseq1/T3;
2. ltpseq1/T7;
3. P21.2/T3;
4. P21.2/T7;
5. P21.1/T3;
6. P21.1/T7;
7. T3/T7.

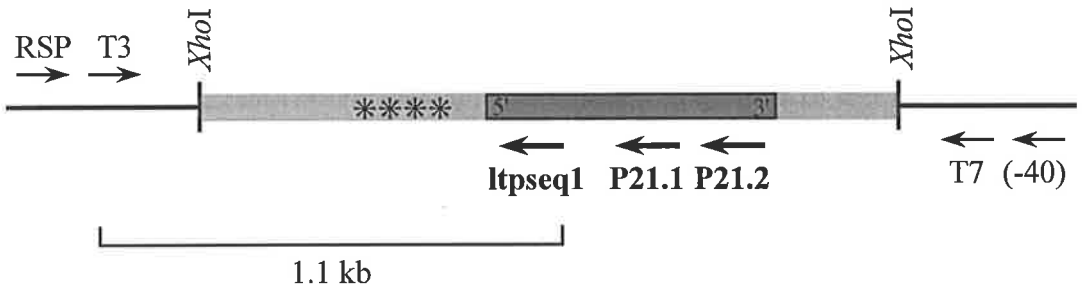
b. Schematic digram of pBluescriptSK(-) subclone pX2.2 (not to scale), showing directions of primers and position of coding (▬) and promoter sequence (✱✱✱✱). Primers designed within the cFSltp sequence are indicated in bold, with primers complementary to pBluescriptSK(-) sequences (——) in normal type.

c. Schematic digram of pBluescriptSK(-) subclone pXE2.7 (not to scale), showing directions of primers and position of coding (▬) and promoter sequence (✱✱✱✱). Primers designed within the cFSltp sequence are indicated in bold, with primers complementary to pBluescriptSK(-) sequences (——) in normal type.

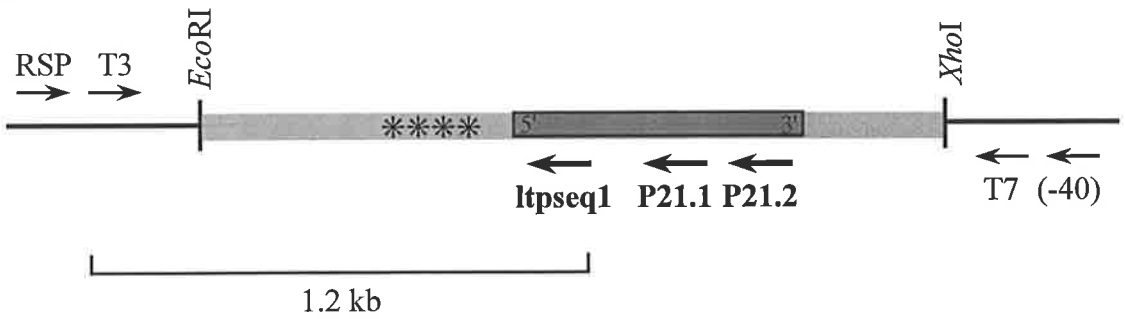

pXE2.7
pX2.2
λFS6(B)
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7

a.

b.


RSP
T3
XhoI
XhoI
5'
3'
ltpseq1
P21.1
P21.2
T7
(-40)
1.1 kb

c.


RSP
T3
EcoRI
XhoI
5'
3'
ltpseq1
P21.1
P21.2
T7
(-40)
1.2 kb

T7 resulted in DNA fragments which corresponded in size to the entire inserts, 2.2 kb and 2.7 kb respectively (Figure 5.10a, lane 7). Minor PCR products of greater size than this are evident in almost all lanes and derive from amplification of the entire plasmid clone. As expected, amplification of λFS6(B) DNA using T3 and T7 primers was negative (Figure 5.10a, lane 7).

Use of primer ltpseq1 in combination with T3 or T7 allowed the orientation of each gene to be deduced and indicated approximately the length of promoter sequence which could be obtained from each subclone. PCR priming of pX2.2 with ltpseq1 and T3 amplified a fragment of 1.1 kb (Figure 5.10a, lane 1) which was absent in a similar reaction utilising T7 instead of T3 (Figure 5.10a, lane 2). Promoter sequences of *FSltp1* therefore lie adjacent to the T3 promoter of pBluescript®SK(-) (Figure 5.10b) and given the position of primer ltpseq1 in the cFSltp sequence (Figure 5.4), the 1.1 kb PCR product contains approximately 800 bp of promoter sequence. Similarly, PCR priming of pXE2.7 with ltpseq1 and T3 amplified a DNA fragment of 1.2 kb (Figure 5.10a, lane 1). From this we can deduce that the promoter region of *FSltp2* also lies adjacent to the T3 promoter of pBluescript®SK(-) and at the *Eco*RI end of pXE2.7 (Figure 5.10c). The 700 bp product (Figure 5.10a, lane 2) observed from ltpseq1/T7 priming of pXE2.7 is due to arbitrary binding of the primer(s) to pXE2.7 sequences. The gene *FSltp2* is therefore oriented 3' to 5' in the λFS6(B) genomic clone.

Use of λFS6(B) DNA as a template in the reactions produced some minor bands but did not result in any significant amplification. The 1.5 kb product, which is present in both P21.2 reactions for all three DNA samples (Figure 5.10a, lanes 3 and 4), is presumed to be a product which results from multiple priming by P21.2. A binding site for P21.2 must be present in the promoter region of each of the LTP genes, located 1.5 kb upstream from the binding site within the coding region. The position of the latter site relative to the first is conserved in both the pX2.2 and pXE2.7 subclones. PCR reactions in which the annealing temperature was increased to 50°C failed to abolish the band (not shown).

PCR of pXE2.7 using primer P21.1 in combination with either T3 or T7 did not result in

amplification of any DNA sequences (Figure 5.10a, lanes 5 and 6). This is consistent with the result that pFS6, which contains P21.1 sequences, does not hybridise to pXE2.7 (Figure 5.7b, lane 7) and with the hypothesis that pXE2.7 contains an LTP gene, *FSltp2*, which is truncated or divergent in sequence at its 3' end. PCR of pX2.2 with P21.1 resulted in products both in combination with T3 and T7 primers (Figure 5.10a, lanes 5 and 6). Priming with P21.1/T3 amplified a product of approximately 1.5 kb and priming with P21.1/T7 amplified a fragment of approximately 700 bp in length. However, since P21.1 and ltpseq1 prime in the same direction (Figure 5.4), a product is only expected using T3. A product size of 1.5 kb is in agreement with the P21.2 and ltpseq1 priming sites separated by a distance of 400 bp (Figure 5.4).

The polarity of fragment X2.2 in λFS6(B) was also deduced as follows. If the 5' end and promoter region of the gene were adjacent to the long arm of λGEM-11™, then PCR products of 1.5 kb and 1.1 kb would be generated in PCR reactions with T7 together with P21.1 and ltpseq1 respectively. Absence of such products (Figure 5.10a, lanes 2 and 6) implies that the promoter region of *FSltp1* is adjacent to the *Xho*I/*Eco*RI fragment of λFS6(B) and that the genes are oriented in the same direction. Confirmation of this orientation came from PCR experiments using primers P21-1 and ltpseq-1 (Table 2.3b), which bind to the coding DNA strand and prime DNA synthesis towards the 3' end of the gene. A 1.1 kb fragment of λFS6(B) was amplified using primers T7 and P21-1 (not shown) and a 1.5 kb fragment was generated using primers T7 and ltpseq-1 (not shown). It can be inferred that the direction of transcription for both genes, if they are transcribed, is from right to left in Figure 5.8. This LTP gene organisation mirrors that in *Sorghum*, where the genes are in tandem and transcribed in the same direction (Pelèse-Siebenbourg *et al.*, 1994). In the genome of tobacco, the two genes, one of which encodes a truncated protein product, are transcribed in opposite directions (Fleming *et al.*, 1992).

The sequences of subclones pX2.2 and pXE2.7 were determined using the above

primers, together with those designed from the cFSltp and pX2.2 sequences (Table 2.3b).

5.8 CHARACTERISATION OF *FSltp1*

5.8.1 The Coding Region of *FSltp1*

Subclone pX2.2 (containing a putative gene designated *FSltp1*) was sequenced in its entirety to yield 2174 bp of sequence (Figure 5.11). An additional 300 bp of *FSltp1* promoter sequence was obtained from the adjoining 2.7 kb *Xho*I/*Eco*RI fragment of λFS6(B) (Figure 5.8). *FSltp1* encodes a putative protein product of 120 amino acids and like FSLTP, the conceptual translation contains features of LTP proteins including presence of a signal peptide at the N-terminus. No other significant ORFs were detected in the sequence of clone pX2.2. *FSltp1* contains an intron of 80 bp which occurs at the C-terminus of the protein, between amino acids 118 and 119 of the conceptual translation, and an insertion of 33 bp was detected in the 3' untranslated region of the gene which is absent from pFS6 group transcripts (Figure 5.11).

The existence of an 80 bp intron in the cotton *FSltp1* gene, which separates two exons encoding 118 and 2 amino acids (Figure 5.11), is in agreement with the presence of introns in other LTP genes. The barley, tobacco, sorghum, broccoli and rice LTP genes are interrupted by an intron in the same position (Linnestad *et al.*, 1991; Fleming *et al.*, 1992; Pelèse-Siebenbourg *et al.*, 1994; Pyee *et al.*, 1994; Vignols *et al.*, 1994). The *FSltp1* intron has the features of plant intron splicing consensus sequences (Brown, 1986), being flanked by the nucleotide sequences 5'-AG/GU-3' and 5'-UGCAG/G-3' at the 5' and 3' junctions respectively. In particular, the GU dinucleotide at the 5' splice site and the AG dinucleotide at the 3' splice site are absolutely conserved and found in all introns, with very rare exceptions. These sequence features are lacking in the 33 bp insertion, which is in the 3' untranslated region of *FSltp1*. Intron recognition in plants is dependant on AU-motifs within introns and presence of such motifs can compensate for suboptimal splice sites, resulting in processing of introns

containing splice sites which deviate strongly from the consensus (Goodall and Filipowicz, 1989; Goodall and Filipowicz, 1991; Luehrsen and Walbot, 1994). However, no AU-tracts are present in the 33 bp insertion of *FSltp1*. In addition, a length of 33 bp is unusually short for a plant intron, which have a minimum length of around 70 bp but can extend for many kb (Goodall and Filipowicz, 1991). Introns shorter than 70 bp do exist (Goodall and Filipowicz, 1991), but it is unknown how efficiently such introns are processed *in vivo*.

All dicot introns characterised thus far are AU-rich and must contain at least 60% AU overall to be recognised and spliced effectively *in vivo* (Wiebauer *et al.*, 1988; Goodall and Filipowicz, 1989; Goodall and Filipowicz, 1991). Both the 80 bp intron and the 33 bp insertion of *FSltp1* satisfy this criterion. The intron is 70% AU, whilst the insertion is 61% AU, compared with 45% for the protein coding region of the gene. However, it seems probable that the 33 bp insertion of *FSltp1* is not an intron but a region of sequence which is present in some members of the LTP gene family and absent in others. If *FSltp1* encodes a fibre-specific mRNA, this 33 bp is spliced out of a primary transcript to give the mature pFS6-group transcript(s) isolated from the cDNA library.

There are two sequence alignments possible between pX2.2 and cFSltp which differ in the regions flanking the 33 bp insertion of *FSltp1*. Firstly, there may be a deletion in the 3' UTR sequence of *FSltp1* compared with the cDNA sequence, such that the surrounding nucleotides are identical between the two sequences. If the single nucleotide deletion is not considered in the alignment, then there are two nucleotide mismatches between the genomic and cDNA sequences in this region. The latter of the two alignments is depicted in Figure 5.12 and has been considered in all sequence comparisons between *FSltp1* and cFSltp.

Like cFSltp, the putative translation initiation codon of *FSltp1* resembles the consensus sequence determined for plant initiation codons, AACAAUGGC (Joshi, 1987b; Lütcke *et al.*, 1987) and matches in five out of nine nucleotide positions. The 3' UTR of *FSltp1* contained three putative polyadenylation signals, located at nucleotide positions 577, 676 and 685 in the

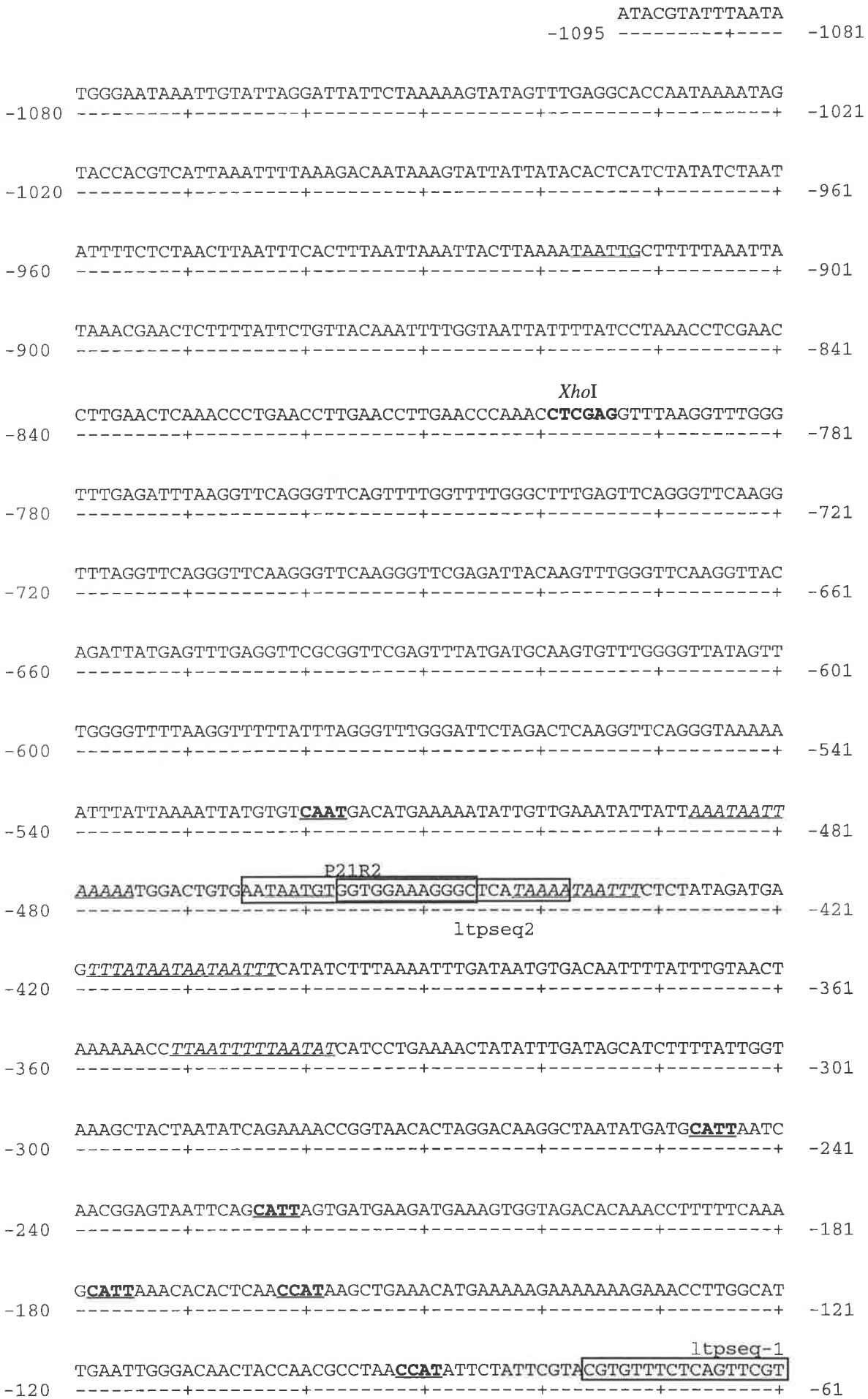
Figure 5.11: Sequence and conceptual translation of a cotton LTP gene, *FSltp1*

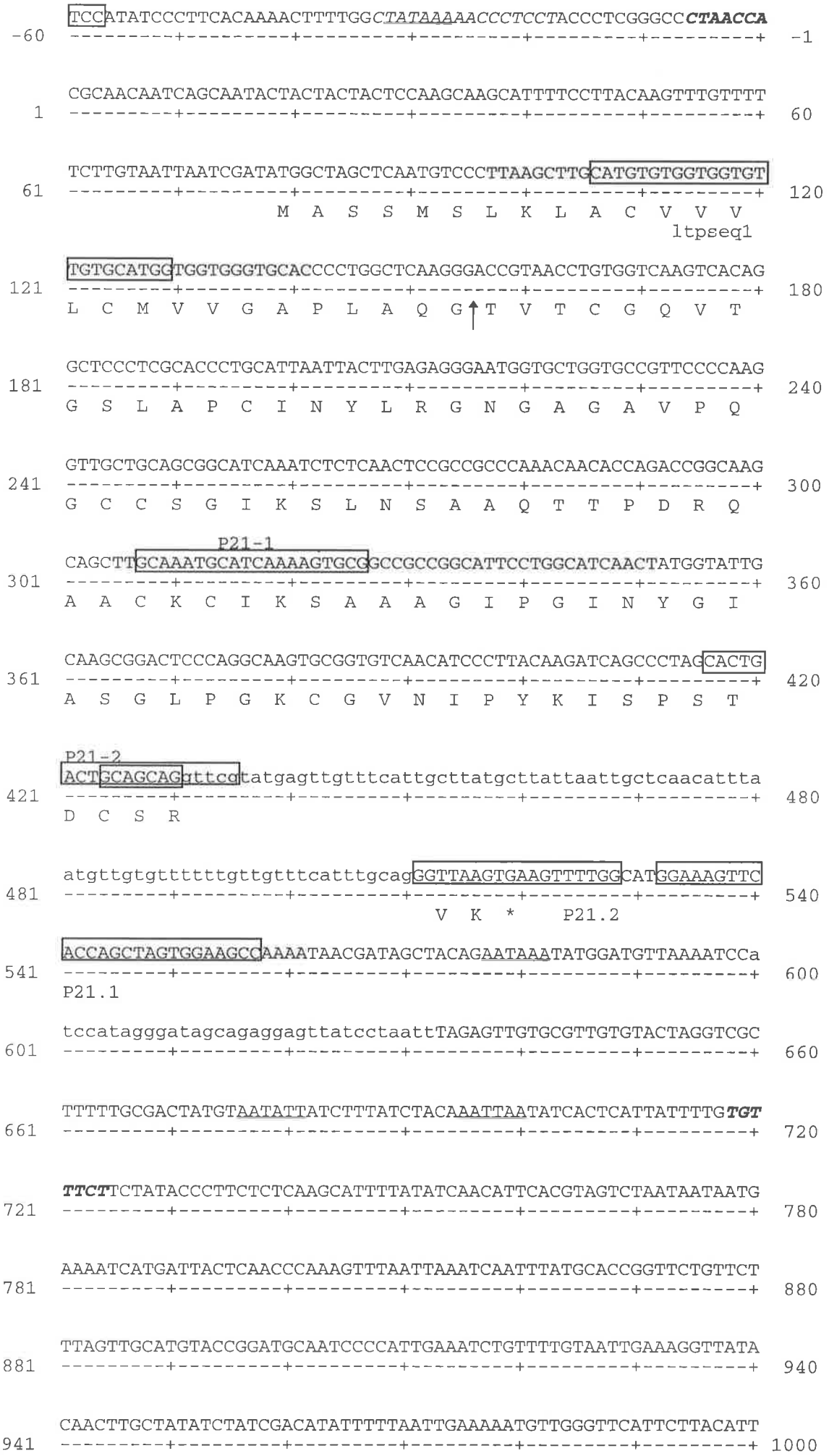
Complete sequence of plasmid subclone pX2.2, combined with 300bp of sequence from the *Xho*I/*Eco*RI flanking fragment of the genomic clone. The *Xho*I site which separates the two subclone sequences is at nucleotide postition -800 to -795 and is in bold.

The putative transcription start site has been designated +1 and the polyadenylation site, as determined by comparison with the cDNA, is at nucleotide position 705. A putative TATA box at nucleotide position -33 to -28 and putative CCAT boxes at positions -521, -248, -179, -163, -143 and -93 in the promoter are in bold and underlined as are three putative polyadenylation signals in the 3' region which identical to those in the cFSltp cDNA. The predicted amino acid sequence of the longest ORF is shown in single letter code below the DNA sequence and both the intron and 33 bp insertion are in lower case lettering. The putative signal sequence cleavage site is indicated by an arrow. A sequence related to the consensus for the start of transcription of plant genes extends from nucleotide position -7 to -1 and is in bold italics, as is a GU-rich cluster from nucleotide position 718 to 724 in the 3' UTR of *FSltp1*.

AT-rich regions found upstream of all the genes encoding LTPs are italicised and underlined. The region adjacent to the TATA box which shows homology to elements in LTP promoters from broccoli, *Arabidopsis* and tomato is also in italics. Imperfect *myb* and *myc* protein binding sites, located at nucleotide positions -918 and -463 respectively, are double underlined.

Primer sequences utilised for sequencing of the subclone are boxed. Those which allowed sequencing of the coding strand of DNA are named below the relevant sequence, while primers which allowed sequencing in the opposite direction (of the mRNA-like strand) are named above the relevant sequence (see also Table 2.3b).


ATACGTATTTAATA
-1095 ---------+---- -1081
TGGGAATAAATTGTATTAGGATTATTCTAAAAAGTATAGTTTGAGGCACCAATAAAATAG
-1080 ---------+---------+---------+---------+---------+---------+ -1021
TACCACGTCATTAAATTTTAAAGACAATAAAGTATTATTATACACTCATCTATATCTAAT
-1020 ---------+---------+---------+---------+---------+---------+ -961
ATTTTCTCTAACTTAATTTCACTTTAATTAAATTACTTAAAATAATTGCTTTTTAAATTA
-960 ---------+---------+---------+---------+---------+---------+ -901
TAAACGAACTCTTTTATTCTGTTACAAATTTTGGTAATTATTTTATCCTAAACCTCGAAC
-900 ---------+---------+---------+---------+---------+---------+ -841
XhoI
CTTGAACTCAAACCCTGAACCTTGAACCTTGAACCCAAACCTCGAGGTTTAAGGTTTGGG
-840 ---------+---------+---------+---------+---------+---------+ -781
TTTGAGATTTAAGGTTCAGGGTTCAGTTTTGGTTTTGGGCTTTGAGTTCAGGGTTCAAGG
-780 ---------+---------+---------+---------+---------+---------+ -721
TTTAGGTTCAGGGTTCAAGGGTTCAAGGGTTCGAGATTACAAGTTTGGGTTCAAGGTTAC
-720 ---------+---------+---------+---------+---------+---------+ -661
AGATTATGAGTTTGAGGTTCGCGGTTCGAGTTTATGATGCAAGTGTTTGGGGTTATAGTT
-660 ---------+---------+---------+---------+---------+---------+ -601
TGGGGTTTTAAGGTTTTTATTTAGGGTTTGGGATTCTAGACTCAAGGTTCAGGGTAAAAA
-600 ---------+---------+---------+---------+---------+---------+ -541
ATTTATTAAAATTATGTGTCAATGACATGAAAAATATTGTTGAAATATTATTAAATAATT
-540 ---------+---------+---------+---------+---------+---------+ -481
P21R2
AAAAATGGACTGTGAATAATGTGGTGGAAAGGGCTCATAAAATAATTTCTCTATAGATGA
-480 ---------+---------+---------+---------+---------+---------+ -421
ltpseq2
GTTTATAATAATAATTTCATATCTTTAAAATTTGATAATGTGACAATTTTATTTGTAACT
-420 ---------+---------+---------+---------+---------+---------+ -361
AAAAAACCTTAATTTTTAATATCATCCTGAAAACTATATTTGATAGCATCTTTTATTGGT
-360 ---------+---------+---------+---------+---------+---------+ -301
AAAGCTACTAATATCAGAAAACCGGTAACACTAGGACAAGGCTAATATGATGCATTAATC
-300 ---------+---------+---------+---------+---------+---------+ -241
AACGGAGTAATTCAGCATTAGTGATGAAGATGAAAGTGGTAGACACAAACCTTTTTCAAA
-240 ---------+---------+---------+---------+---------+---------+ -181
GCATTAAACACACTCAACCATAAGCTGAAACATGAAAAAGAAAAAAAGAAACCTTGGCAT
-180 ---------+---------+---------+---------+---------+---------+ -121
ltpseq-1
TGAATTGGGACAACTACCAACGCCTAACCATATTCTATTCGTACGTGTTTCTCAGTTCGT
-120 ---------+---------+---------+---------+---------+---------+ -61

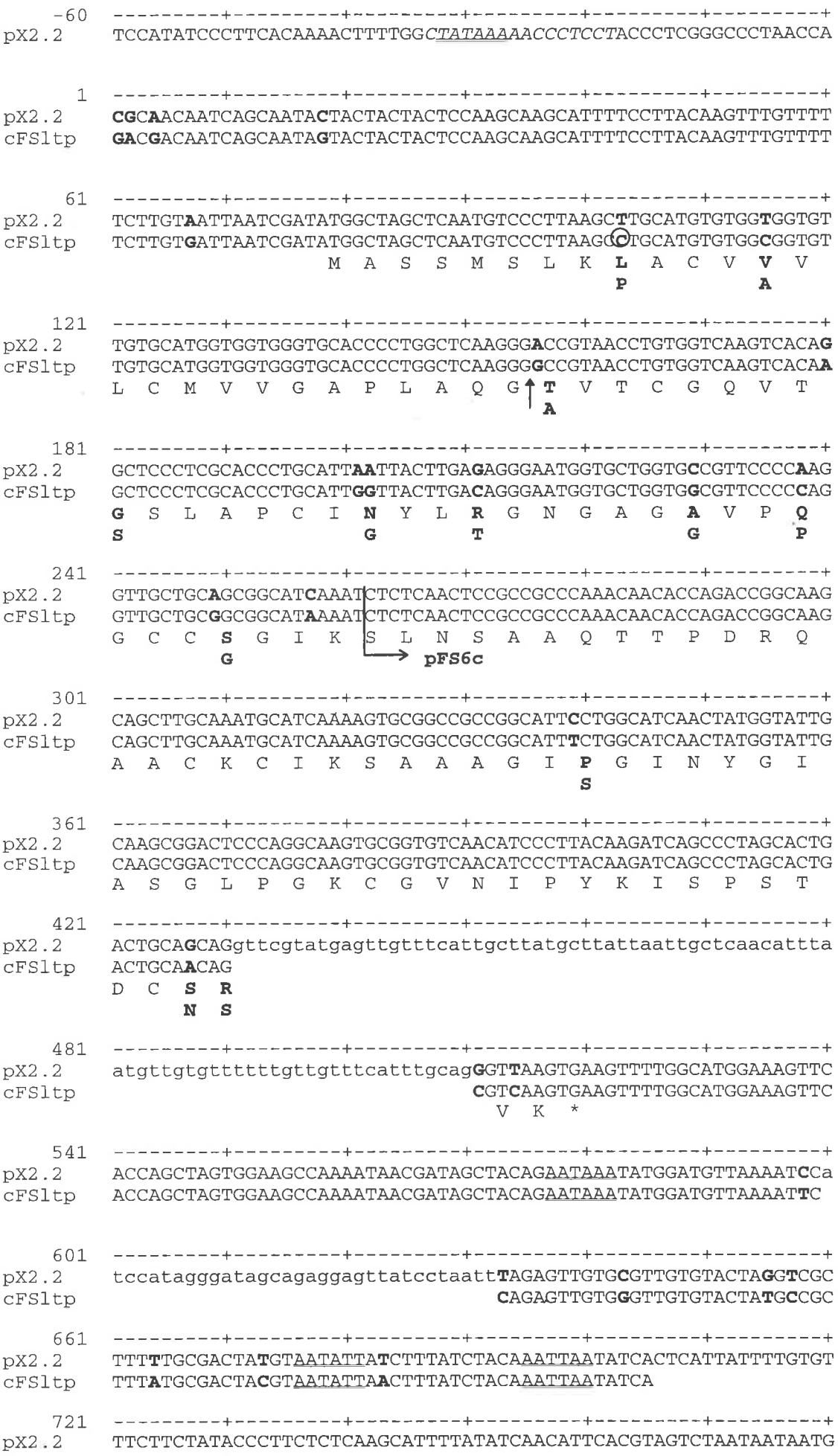

TCCATATCCCTTCACAAAACTTTTGGCTATAAAAACCCTCCTACCCTCGGGCCCTAACCA
-60
-1
CGCAACAATCAGCAATACTACTACTACTCCAAGCAAGCATTTTCCTTACAAGTTTGTTTT
1
60
TCTTGTAATTAATCGATATGGCTAGCTCAATGTCCCTTAAGCTTGCATGTGTGGTGGTGT
61
120
M A S S M S L K L A C V V V
ltpseq1
TGTGCATGGTGGTGGGTGCACCCCTGGCTCAAGGGACCGTAACCTGTGGTCAAGTCACAG
121
180
L C M V V G A P L A Q G T V T C G Q V T
GCTCCCTCGCACCCTGCATTAATTACTTGAGAGGGAATGGTGCTGGTGCCGTTCCCCAAG
181
240
G S L A P C I N Y L R G N G A G A V P Q
GTTGCTGCAGCGGCATCAAATCTCTCAACTCCGCCGCCCAAACAACACCAGACCGGCAAG
241
300
G C C S G I K S L N S A A Q T T P D R Q
P21-1
CAGCTTGCAAATGCATCAAAAGTGCGGCCGCCGGCATTCCTGGCATCAACTATGGTATTG
301
360
A A C K C I K S A A A G I P G I N Y G I
CAAGCGGACTCCCAGGCAAGTGCGGTGTCAACATCCCTTACAAGATCAGCCCTAGCACTG
361
420
A S G L P G K C G V N I P Y K I S P S T
P21-2
ACTGCAGCAGgttcgtatgagttgtttcattgcttatgcttattaattgctcaacattta
421
480
D C S R
atgttgtgtttttttgttgtttcatttgcagGGTTAAGTGAAGTTTTGGCATGGAAAGTTC
481
540
V K * P21.2
ACCAGCTAGTGGAAGCCAAAATAACGATAGCTACAGAATAAATATGGATGTTAAAATCCa
541
600
P21.1
tccatagggatagcagaggagttatcctaattTAGAGTTGTGCGTTGTGTACTAGGTCGC
601
660
TTTTTGCGACTATGTAATATTATCTTTATCTACAAATTAATATCACTCATTATTTTGTGT
661
720
TTCTTCTATACCCTTCTCTCAAGCATTTTATATCAACATTCACGTAGTCTAATAATAATG
721
780
AAAATCATGATTACTCAACCCAAAGTTTAATTAAATCAATTTATGCACCGGTTCTGTTCT
781
880
TTAGTTGCATGTACCGGATGCAATCCCCATTGAAATCTGTTTTGTAATTGAAAGGTTATA
881
940
CAACTTGCTATATCTATCGACATATTTTTAATTGAAAAATGTTGGGTTCATTCTTACATT
941
1000


ACTTTAAATCTTTGAAGTGGACGACTCTTTTGGAGACATTATTATATGTGATTGTGTTTC
1001 ---------+---------+---------+---------+---------+---------+ 1060
P21F2
TTTTTAAATGAAAATATCAAGTTTTATAAAAATATAATTAAGAGAGATTATTCATCTGAT
1061 ---------+---------+---------+---------+---------+---------+ 1120
CATTTTGAGATGTGGATCAAAGTTCATGATCATTGTGCATAGACGTTTTATTGAGATTAA
1121 ---------+---------+---------+---------+---------+---------+ 1180
CTAATTAAATTGGGCTCTATGATCCGCTTGAACTTGTCCAACTCGTGGAACATATGGAGA
1181 ---------+---------+---------+---------+---------+---------+ 1240
AGCTCATCTACTTAAACCTTTGGTGGTTCAGTGAAACTCTCCAGTAAAACTAGAGTTACC
1241 ---------+---------+---------+---------+---------+---------+ 1300
AGAGTAATTAATGTAAATCCTAAGGGATAAAGATGTATAGAGTTTAAAACCCTTAAGACT
1301 ---------+---------+---------+---------+---------+---------+ 1360
CATCTTGTAGATAAGTACTAATCTCAATCGTTGATGTAATTCAATCTATACTATTGGATC
1361 ---------+---------+---------+---------+---------+---------+ 1420

Figure 5.12: Sequence comparison between a fibre-specific LTP cDNA and a cotton LTP gene, *FSltp1*

Sequence comparison between the pFS6 consensus sequence and the relevant region of genomic subclone pX2.2. The 28 single nucleotide substitutions are shown in bold type and both the intron and 33 bp insertion of *FSltp1* are in lower case lettering. The predicted amino acid sequence of the ORF encoded by *FSltp1* is shown in single letter code below the DNA sequences and the 14 amino acid changes are indicated in bold type. The putative signal sequence cleavage site is indicated by an arrow.

All other sequence features are marked as in Figure 6.11.


-60
-1
pX2.2 TCCATATCCCTTCACAAAACTTTTGGCTATAAAAACCCTCCTACCCTCGGGCCCTAACCA
1
60
pX2.2 CGCAACAATCAGCAATACTACTACTACTCCAAGCAAGCATTTTCCTTACAAGTTTGTTTT
cFSltp GACGACAATCAGCAATAGTACTACTACTCCAAGCAAGCATTTTCCTTACAAGTTTGTTTT
61
120
pX2.2 TCTTGTAATTAATCGATATGGCTAGCTCAATGTCCCTTAAGCTTGCATGTGTGGTGGTGT
cFSltp TCTTGTGATTAATCGATATGGCTAGCTCAATGTCCCTTAAGCCTGCATGTGTGGCGGTGT
M A S S M S L K L A C V V V
P A
121
180
pX2.2 TGTGCATGGTGGTGGGTGCACCCCTGGCTCAAGGGACCGTAACCTGTGGTCAAGTCACAG
cFSltp TGTGCATGGTGGTGGGTGCACCCCTGGCTCAAGGGGCCGTAACCTGTGGTCAAGTCACAA
L C M V V G A P L A Q G T V T C G Q V T
A
181
240
pX2.2 GCTCCCTCGCACCCTGCATTAATTACTTGAGAGGGAATGGTGCTGGTGCCGTTCCCCAAG
cFSltp GCTCCCTCGCACCCTGCATTGGTTACTTGACAGGGAATGGTGCTGGTGGCGTTCCCCCAG
G S L A P C I N Y L R G N G A G A V P Q
S G T G P
241
300
pX2.2 GTTGCTGCAGCGGCATCAAATCTCTCAACTCCGCCGCCCAAACAACACCAGACCGGCAAG
cFSltp GTTGCTGCGGCGGCATAAAATCTCTCAACTCCGCCGCCCAAACAACACCAGACCGGCAAG
G C C S G I K S L N S A A Q T T P D R Q
G
pFS6c
301
360
pX2.2 CAGCTTGCAAATGCATCAAAAGTGCGGCCGCCGGCATTCCTGGCATCAACTATGGTATTG
cFSltp CAGCTTGCAAATGCATCAAAAGTGCGGCCGCCGGCATTTCTGGCATCAACTATGGTATTG
A A C K C I K S A A A G I P G I N Y G I
S
361
420
pX2.2 CAAGCGGACTCCCAGGCAAGTGCGGTGTCAACATCCCTTACAAGATCAGCCCTAGCACTG
cFSltp CAAGCGGACTCCCAGGCAAGTGCGGTGTCAACATCCCTTACAAGATCAGCCCTAGCACTG
A S G L P G K C G V N I P Y K I S P S T
421
480
pX2.2 ACTGCAGCAGgttcgtatgagttgtttcattgcttatgcttattaattgctcaacattta
cFSltp ACTGCAACAG
D C S R
N S
481
540
pX2.2 atgttgtgtttttttgttgtttcatttgcagGGTTAAGTGAAGTTTTGGCATGGAAAGTTC
cFSltp CGTCAAGTGAAGTTTTGGCATGGAAAGTTC
V K *
541
600
pX2.2 ACCAGCTAGTGGAAGCCAAAATAACGATAGCTACAGAATAAATATGGATGTTAAAATCCa
cFSltp ACCAGCTAGTGGAAGCCAAAATAACGATAGCTACAGAATAAATATGGATGTTAAAATTC
601
660
pX2.2 tccatagggatagcagaggagttatcctaatttTAGAGTTGTGCGTTGTGTACTAGGTCGC
cFSltp CAGAGTTGTGGGTTGTGTACTATGCCGC
661
720
pX2.2 TTTTTGCGACTATGTAATATTATCTTTATCTACAAATTAATATCACTCATTATTTTGTGT
cFSltp TTTATGCGACTACGTAATATTAACTTTATCTACAAATTAATATCA
721
780
pX2.2 TTCTTCTATACCCTTCTCTCAAGCATTTTATATCAACATTCACGTAGTCTAATAATAATG

pX2.2 sequence (Figure 5.11). The signals are identical to those detected in the cFSltp cDNA sequence and all three may be functional *in vivo*. A GU-rich sequence located 13 bp downstream of the putative polyadenylation site (Figure 5.11) is reminiscent of a well-defined *cis*-acting sequence found commonly in mammalian genes which, together with the polyadenylation signal AATAAA, controls mRNA 3'-end processing (Keller, 1995). A similar GU-like cluster with consensus sequence YGUGUUYY (where Y is a pyrimidine) has been identified downstream of poly(A) sites in several plant genes (Joshi, 1987a).

There are 28 nucleotide differences between the cDNA sequence cFSltp and the sequence of gene *FSltp1* (Figure 5.12), which result in only 95% similarity between the fibre-specific mRNA and the candidate gene. Of the 28 nucleotide substitutions, 16 occur in the 261 bp portion of cFSltp sequence which is derived from the 5'-RACE clone (Figure 5.4; Figure 5.12), so that 94% identity exists between the mRNA and the 5' section of the gene. This suggests that reverse transcriptase or *Taq* DNA polymerase errors account for a proportion of the discrepancies between the cDNA and the gene. A misincorporation during the early rounds of amplification might appear in every clone derived from that particular product, which may be the case for the pFS6 5'-RACE clones. One of the differences corresponds to the single nucleotide mismatch between the two RACE clones and may be attributed to a *Taq* DNA polymerase error during PCR. The *FSltp1* sequence matches that of pFS6-56, having a T at nucleotide position 104, instead of the C present in RACE clone pFS6-30. The remaining 12 nucleotide changes are confined to the 331 bp sequence derived from the pFS6 consensus cDNA sequence (Figure 5.12) and these are clustered near the intron-exon boundaries, resulting in 98% nucleotide identity for the 3' section of the gene. Furthermore, the protein-coding region of the genomic clone contains 15 of the 26 nucleotide substitutions, resulting in 12 amino acid substitutions (Figure 5.12) and only 90% amino acid identity between the open reading frames of the cDNA and the candidate gene.

The nucleotide differences, together with the 33 bp insertion in the 3' UTR of *FSltp1*,

suggest that the gene which encodes the pFS6 group fibre-specific LTP transcripts has yet to be isolated. In addition, *FSltp1* is even less similar in nucleotide sequence to GH3 (Ma *et al.*, 1995) than to cFSltp such that *FSltp1* is unlikely to encode either of the two cotton fibre-specific LTPs. Notably, *FSltp1* cannot encode the longest class of pFS6 transcript, represented by cDNA clones pFS8 and pFS6 (Figure 5.1; Table 5.1). If the extended length of pFS8 and pFS19 is not due to cloning artefacts, then they must be encoded by an LTP gene other than *FSltp1*.

5.8.2 The Promoter of a Cotton LTP Gene, *FSltp1*

Typically, the regulatory elements that are responsible for directing specific transcription initiation reside within the 5' promoter region of a gene. Characterisation of promoter regions is therefore crucial to gaining an understanding of the basis for differing LTP gene expression patterns and, more pertinent to this study, for fibre-specific expression patterns. The 5' flanking region of *FSltp1* contains a 5'-TATAAA-3' motif, located at nucleotide positions -33 to -28 relative to the putative transcription start site (Figure 5.11) that resembles the Goldberg-Hogness TATA box (Efstratiadis *et al.*, 1980). The presence of TATA sequences in mammalian genes is indicative of an inducible gene versus a housekeeping gene (Dynan, 1986). However, most of the plant genes studied thus far are inducible, making a similar comparison impossible. Upstream from the putative cap site are several CCAT and CATT motifs that may be related to the CAAT box sequence 5'-GG(C/T)CAATCT-3' (Shenk, 1981). However, the sequence similarity in each case is poor and it may be that the *FSltp1* gene does not contain a CAAT-like sequence, an observation not uncommon among plant genes (McClure *et al.*, 1989).

The putative transcription start site of *FSltp1* was determined by comparison between genomic and cDNA sequences and its position corroborated by the presence of several features common to plant genes. Partial homology to the consensus sequence for the start of transcription, 5'-CTCATCA-3' (Joshi, 1987b), was found 20 bp downstream of the TATA

box, at nucleotide position -7 in the pX2.2 sequence (Figure 5.11). The sequence is adjacent to the putative start of *FSltp1* transcription as determined by comparison with the cDNA clone sequence. Joshi (1987b) showed that the length of the leader sequence in a compilation of 73 plant genes varied from 9 nt to 193 nt, with a predominant length of 40-80 nts, and the leader sequences were generally AT-rich. Correspondingly, the putative leader sequence of *FSltp1* is 77 bp in length and consists of 66% AT compared with an AT content of 45% for the coding region of the gene. In addition, the mean distance of 32±7 nts between the putative TATA box and transcription start site (Joshi, 1987b) is maintained for *FSltp1* which has 29 bp in this region (Figure 5.11).

Approximately 1 kb of the 5' non-coding region of *FSltp1* was sequenced and found to be AT-rich, consisting of ~~only~~ only 68% AT compared with 45% AT for the protein-coding region of *FSltp1*. Accordingly, the promoter sequence shows significant homology to telomeric repeats and mitochondrial sequences in the nucleotide databases (GenBank and EMBL; January, 1996), particularly over the first 700 bp of the *FSltp1* promoter. AT-rich regions have been identified in several plant LTP promoters including *FSltp1* (Linnestad *et al.*, 1991; Vignols *et al.*, 1994; Figure 5.11) and may constitute a novel class of *cis*-acting regulatory element which is present in a diverse set of plant genes (reviewed in Forde, 1994). The region adjacent to the TATA box (Figure 5.11) shows homology to region IV identified in the broccoli LTP genes (Pyee and Kolattukudy, 1995), which is also present in corresponding *Arabidopsis* and tobacco LTP promoter sequences (Fleming *et al.*, 1992).

Putative regulatory elements involved in LTP gene expression have been identified only from barley LTP genes. Short nucleotide sequences with homology to *myb* and *myc* protein binding sites have been detected upstream of the barley *Ltp1* and *Ltp2* genes and have been suggested to be involved in the regulation of their expression (Linnestad *et al.*, 1991; Kalla *et al.*, 1994). Although two related sequences were identified in the cotton *FSltp1* promoter (Figure 5.11), they present differences with the consensus sequences of these sites and such

modifications are usually sufficient to disrupt the efficiency of the site (Roth *et al.*, 1991). Another sequence motif present in the barley promoter is common to several genes expressed in aleurone layers and may be a site for activator protein binding (Linnestad *et al.*, 1991).

Generally however, few regions were found in *FSltp1* which show similarity to other plant *Ltp* promoters. Given that LTP genes exhibit such differing expression patterns, it seems reasonable that they will be subject to distinct regulatory mechanisms. A comparison between the promoter regions of cotton fibre-specific genes is required in order to determine whether there is a common *cis* regulatory element that functions to confer the differential gene expression. Several fibre-specific genes have recently been reported (Delmer *et al.*, 1995; John, 1995; John and Crow, 1992; John and Keller, 1995; Ma *et al.*, 1995), but identification of possible control regions in cotton fibre-specific promoters is hampered by a lack of sequence information. Promoter regions of cotton fibre-specific genes have been shown to direct tissue-specific activity of GUS in both transient expression studies and in transgenic cotton plants (John, 1995; John and Crow, 1992).

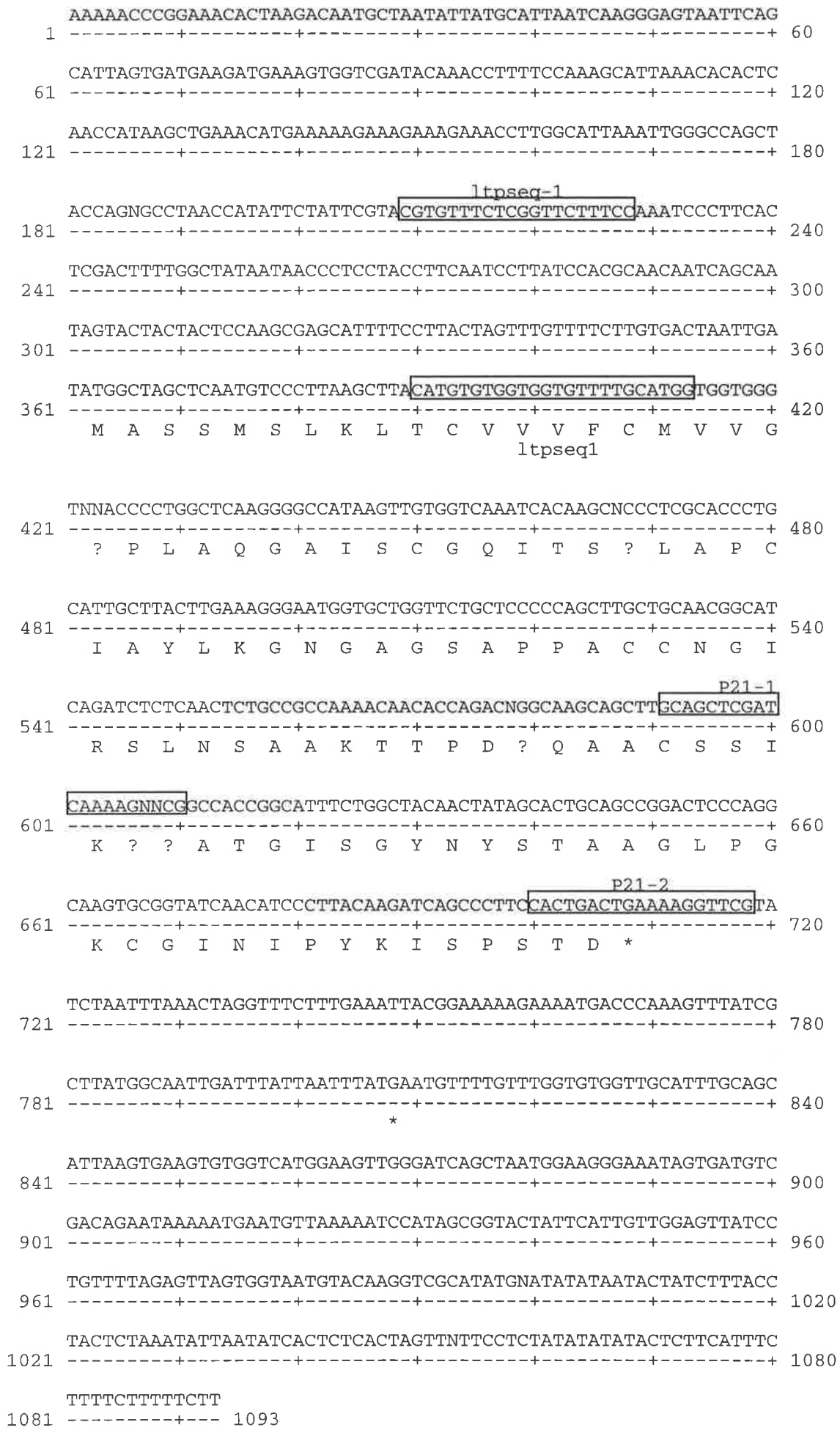
5.9 Sequence Analysis of *FSltp2*

Partial sequencing of the pXE2.7 subclone (containing a second putative gene designated *FSltp2*) utilised primers ltpseq1, ltpseq-1, P21-1 and P21-2 and yielded 1092 bp of sequence which was derived mostly from one DNA strand only (Figure 5.13). The sequence contained an open reading frame which was similar to both FSLTP and the conceptual translation of *FSltp1* but which contained a stop codon after amino acid 115 of putative protein, and may therefore encode a truncated LTP product. A sequencing error in the vicinity of this putative stop codon cannot be ruled out. The 348 bp coding sequence of *FSltp2* was different from the consensus cDNA sequence cFSltp, with at most 89% nucleotide identity, and the 24 amino acid substitutions in the 115 residue protein results in only 78% identity between FSLTP and the conceptual translation of *FSltp2*. The low degree of similarity to the cDNA and the

Figure 5.13: Sequence and conceptual translation of cotton LTP gene *FSltp2*

Partial nucleotide sequence of genomic subclone pXE2.7, with the conceptual translation of a truncated LTP shown in single letter code below the DNA sequence and the putative TGA stop codon indicated by an asterisk. The polyadenylation site as determined by comparison with the cDNA is at nucleotide position 1040. The TGA termination codon which corresponds with cFSltp and *FSltp1* is at nucleotide position 807 and is also indicated by an asterisk.

A putative TATA box at nucleotide position 254 to 259 in the promoter is in bold and underlined. Primer sequences utilised for sequencing the internal regions of the subclone are boxed. Those which allowed sequencing of the mRNA-like strand of DNA are named above the relevant sequence, while primers which allowed sequencing in the opposite direction (of the coding DNA strand) are named below the relevant sequence (see also Table 2.3b).


ltpseq-1
ltpseq1
P21-1
P21-2

truncated protein product rule out *FSltp2* as a gene which encodes any of the fibre-specific cDNAs isolated.

FSltp2, whilst it encodes a truncated protein, has some of the features of *FSltp1*. The two genes can be aligned in the coding region and show 89% sequence identity. However, from the stop codon of *FSltp2* onwards the sequence similarity ends and an alignment shows stretches of homology intermixed with long fragments having no homology (not shown). The 3' region of the two genes have an overall sequence similarity of approximately 70%. This is consistent with the results of Southern and PCR experiments (Section 5.6; Section 5.7) which suggested that *FSltp2* is truncated or diverged in sequence at its 3' end. The sequence immediately upstream of the putative initiation codon ATG is more highly conserved between genes *FSltp1* and *FSltp2*, with a nucleotide similarity of 90%, but *FSltp2* has a slightly modified TATA box (Figure 5.13). Since promoter sequences near the TATA box are often essential for gene activation, these changes in *FSltp2* may render the gene inactive. It is not known whether either of the LTP genes identified in this study is transcribed. A schematic diagram which summarises the organisation of *FSltp1* and *FSltp2* on the λFS6(B) genomic fragment is shown in Figure 5.14.

The genomic DNA fragment which contains *FSltp2* is not hybridised by cDNA clone pFS6 (Figure 5.7b, lane 7) and results summarised above indicate that the sequence of *FSltp2* deviates from that of cFSltp. Presumably then, the 650 nt LTP transcript detected in fibre tissue (Figure 4.4a) is not derived from gene *FSltp2*, since the probe used to detect the mRNA was pFS6. Use of pFS2 as a probe to a Northern blot did not detect any additional transcripts and produced a result identical to that of pFS6 (not shown). It may be that *FSltp2* is a pseudogene or an inactive LTP counterpart or is expressed at low levels in fibre, leaf, flower and/or seedling tissue or in cotton tissues which have not been examined.

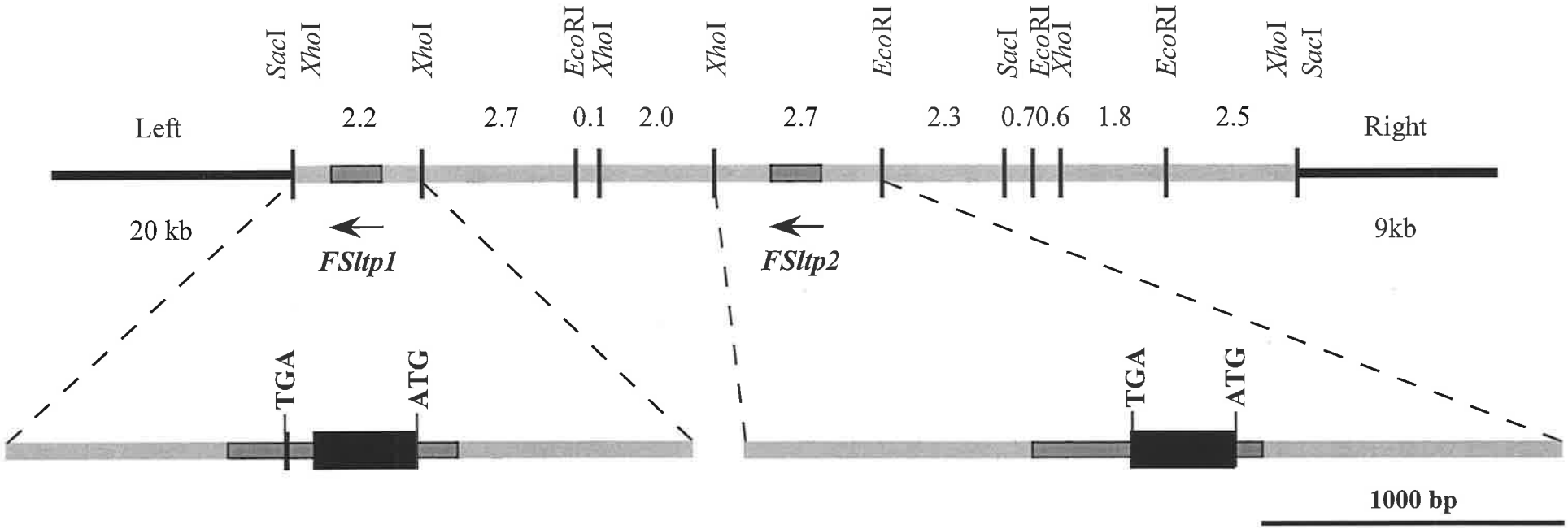


Figure 5.14: Schematic representation of genomic clone λFS6(B)

Gene *FSltp2* is located upstream (right in the figure) from *FSltp1*. Arrows indicate direction of transcription and fragment sizes are indicated in kb. Scale bar refers to the expanded DNA fragments, in which coding regions are shaded ( ) and protein-coding regions are filled ( ).

5.10 Southern Analysis of Fibre-specific cDNA Clone pFS6

The cFSltp transcript, whilst being abundant in fibre tissue, is also present at low levels in leaf, whole flower and seedling tissue (Figure 4.4a). This may be indicative of a multigene family in cotton, in which the different members are differentially expressed or encode divergent mRNA species. Southern analysis using cFSltp as a probe to cotton genomic DNAs identified at least six hybridising genomic DNA fragments (Figure 5.15, lanes 1-8), which indicates that LTPs are encoded in *G. hirsutum* cv. Siokra 1-2 by a gene family. The complexity of the gene family in allotetraploid cotton is a consequence of the presence of two distinct subgenomes. Genes derived from both the *A* and *D* subgenomes are present. In addition, evolution of polymorphic sites may also have occurred since hybridisation occurred 1-2 million years ago (Wendel, 1989). This result contrasts with that of Ma *et al.* (1995), who suggest that the cotton LTP gene is present in one or few copies in *G. hirsutum* cv. DES119. Small LTP gene families have been detected in a variety of plants, and multiple LTP genes have been isolated from the genomes of barley (Linnestad *et al.*, 1991), tobacco (Fleming *et al.*, 1992), sorghum (Pelèse-Siebenbourg *et al.*, 1994) and broccoli (Pyee and Kolattukudy, 1995).

Restriction fragments which corresponded in size to the LTP sequences in the genomic clone λFS6(B) could be identified in the Siokra 1-2 genome. The 4.8 kb *Eco*RI fragment of λFS6(B) was evident (Figure 5.15a, lane 1) as was the 12 kb *Sac*I fragment (Figure 5.15a, lane 8). An identical hybridisation pattern was obtained in other tetraploid cotton varieties, *G. hirsutum* cultivars Siokra 1-4, Siokra L-22, CS7S, Strumica 5086 and Paymaster 909, as well as another species, *G barbadense* cv. Pima S-6 (not shown), which suggests that the LTP gene family is highly conserved in the tetraploid members of the genus *Gossypium*. cFSltp homologues were also detected in the Australian diploid species *G. robinsonii* and *G. sturtianum* (Figure 5.15a; lanes 9-10), but only weak hybridisation occurred to sequences in other plant genomes tested such as carrot and *Arabidopsis* (not shown). This is not surprising

since the cFSltp sequence shows at most 63% nucleotide identity (in the region which encodes the mature polypeptide) to homologues from other plants (Bernhard *et al.*, 1991; Section 5.4).

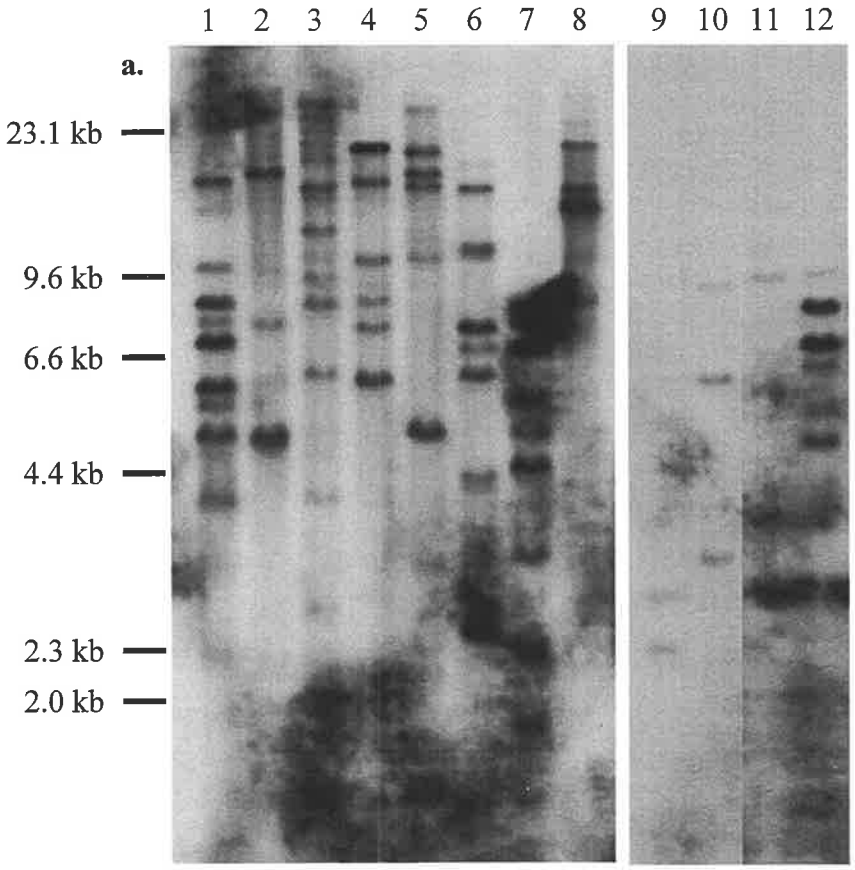
Gossypium hirsutum is a 1-2 million year old allotetraploid derived from two diploid ancestors. *G. herbaceum* and *G. raimondii* are the two diploid species with genomes closest to the *A* and *D* subgenomes of the allotetraploids (reviewed in Endrizzi *et al.*, 1985). Probing of restricted genomic DNA from *G. raimondii* with cFSltp produced several hybridising DNA fragments (Figure 5.15a, lane 12), which suggests that the *G. raimondii* genome contains several LTP-related genes. The hybridisation pattern was identical in DNA samples from two *G. raimondii* plants (not shown) and some of the hybridising fragments were similar in mobility to those detected in *G. hirsutum*. In particular, the three *Eco*RI fragments of *G. raimondii* which hybridised most strongly to the probe (Figure 5.15a, lane 12) appear to be also present in the genome of *G. hirsutum* cv. Siokra 1-2 (Figure 5.15a, lane 1). The cFSltp probe hybridised weakly to DNA fragments in the diploid genome of *G. herbaceum* (Figure 5.15a, lane 11), with a restriction pattern which differed from that of *G. raimondii* (Figure 5.15a, lane 12). Some of the bands detected in *G. herbaceum* appear also to be present in the tetraploid genome of *G. hirsutum*.

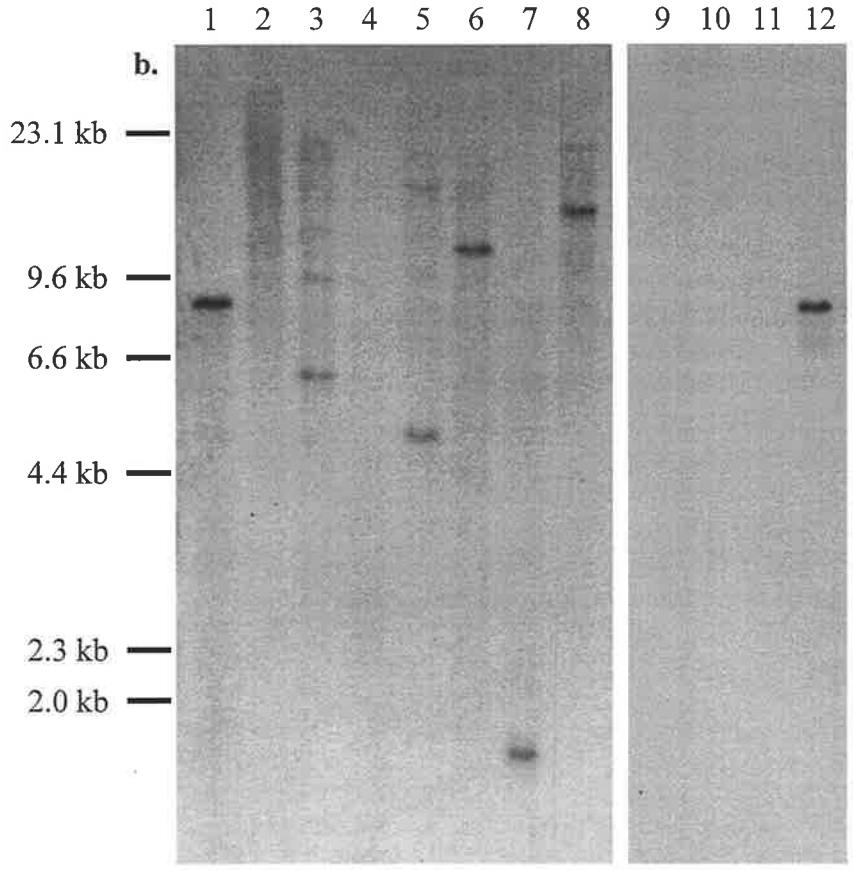
A 2.8 kb *Xho*I fragment of λFS6(B), designated X2.8, was also used to probe the Southern blot of cotton genomic DNAs. The probe is specific for the *Eco*RI fragment which contains *FSltp1* and should detect sequences which are adjacent to *FSltp1* (Figure 5.14), to the exclusion of *FSltp2* sequences. Initial probing with X2.8 detected many regions of homology in all cotton genomes, resulting in a smear of hybridisation in each lane of DNA, upon which were superimposed discrete bands of hybridisation (not shown). Again, the pattern of hybridisation appeared identical for all the *G. hirsutum* varieties and for the single *G. barbadense* variety. Some of the sequences contained within the 2.8 kb *Xho*I fragment of λFS6(B) are therefore repeated and probably dispersed throughout the cotton genome. A wash at high stringency removed much of the bound X2.8 probe, with the retention of a single band

Figure 5.15: Genomic organisation of cotton LTP genes

a. Genomic Southern blot probed with cFSltp sequences, derived from clones pFS6 and pFS6-30. Each lane was loaded with 10 µg of total genomic DNA cut with enzymes that do not cut within the cFSltp sequence. Numbers on the left represent molecular weights of wild type Lambda DNA restricted with *Hin*dIII (kb). Lanes 1-8 contain DNA from *G. hirsutum* L., cv. Siokra 1-2 restricted with: 1. *Eco*RI , 2. *Bam*HI, 3. *Bgl*II, 4. *Eco*RV, 5. *Nco*I, 6. *Xba*I, 7. *Hin*dIII and 8. *Sac*I. Lanes 9, 10, 11 and 12 contain samples of DNA from *G. sturtianum*, *G. robinsonii*, *G. herbaceum* and *G. raimondii* respectively, restricted with *Eco*RI.

b. Genomic Southern blot from part a. probed with the gene-specific probe X2.8, derived from clone λFS6(B). Lanes are labelled as above.


1 2 3 4 5 6 7 8 9 10 11 12
a.
23.1 kb
9.6 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb


1 2 3 4 5 6 7 8 9 10 11 12
b.
23.1 kb
9.6 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb

of hybridisation per restriction enzyme (Figure 5.15b, lanes 1-8), each of which corresponded to a particular band on the Southern probed with the entire LTP cDNA clone (Figure 5.15a, lane 8).

The *FSltp1* sequences are confined to an *Eco*RI fragment of approximately 8 kb in Siokra 1-2 DNA, which is also present in the *G. raimondii* genome (Figure 5.15b, lane 12) but absent from the genome of *G. herbaceum* (Figure 5.15b, lane 11). Taken together, these results suggest that the *FSltp1* gene originated from an ancestral gene present in the diploid *D* genome of *G. raimondii*.

5.11 Summary and Conclusions

A full-length cDNA clone, which corresponds to a gene expressed preferentially in cotton (*Gossypium hirsutum*) fibre cells, has been isolated and characterised. The nucleotide sequence includes an open reading frame of 120 amino acids, encoding a polypeptide which is a member of a conserved group of plant lipid transfer proteins (LTPs). The clone isolated from Siokra 1-2 and described in this Chapter defines at least one new class of fibre-specific LTP and differs significantly from the LTP previously isolated from *G. hirsutum* cv. DES119. Southern hybridisation indicated that LTPs are encoded in *G. hirsutum* cv. Siokra 1-2 by at least six genes which code for similar mRNAs and that the LTP gene family is highly conserved in two members of the genus *Gossypium*. LTP genes were also present in the diploid ancestors of the tetraploid *G. hirsutum*. Most of the hybridising fragments were common to *G. hirsutum* cv. Siokra 1-2 and *G. raimondii*, suggesting that the majority of LTP homologues in the tetraploid genome of *G. hirsutum* are derived from the genome of *G. raimondii*.

The results from analysis of a genomic clone, λFS6(B), confirm that at least two cotton LTP genes are present in the same genomic region, with the two open reading frames separated by a distance of about 5 kb. Close linkage of genes *FSltp1* and *FSltp2* suggests that

the LTP genes arose by gene duplication and divergence, similar to many other gene families in higher plants.

Sequencing of *FSltp1* revealed an open reading frame similar, but not identical to, the cFSltp cDNA. The genomic sequence contained a 3' insertion which is absent from the fibre-specific LTP cDNAs and lacks the conserved features of introns, making it unlikely to be spliced from an *FSltp1* primary transcript. *FSltp1* encodes an appropriate translational start site, stop codon, polyadenylation signal(s) and a putative TATA box, together with some regions similar to the promoter sequences of other plant LTP genes. The sequence divergence between cFSltp and *FSltp1* and presence of a putative 33 bp insertion in the 3' untranslated region of *FSltp1* may indicate that the gene which encodes cFSltp has yet to be isolated. The amino acid sequence corresponding to *FSltp2* appears to encode a truncated LTP of 115 amino acids and cDNAs corresponding to *FSltp2* were not detected in the cDNA library.

CHAPTER 6

Control of Ribosomal RNA Synthesis in the Cotton Fibre

6.1 INTRODUCTION

The processes of cell expansion and cell division require large increases in the synthesis of specific proteins and consequently, may require an increase in the amount of ribosomes and ribosomal RNA (rRNA) per cell. Substantial cell elongation occurs during the development of a cotton fibre and considerable synthesis of proteins occurs during fibre development (Huwyler *et al.*, 1979). More numerous ribosomes and rough endoplasmic reticulum observed in the fibre cells of cotton than in adjacent non-differentiated epidermis (Ramsey and Berlin, 1976a) confirms an increased capacity for protein synthesis. Nucleoli, the sites of rRNA gene transcription and transcript processing, are prominent early in fibre cell development and their apparent disappearance at 20 days after anthesis (Peeters *et al.*, 1988) suggests that the amount of ribosomes synthesised in the early stages of cotton fibre elongation may determine the subsequent rate of cell elongation. In addition, the size of the nucleolus at early stages of fibre development has been shown to be associated with the final length of the cotton fibre (Peeters *et al.*, 1987; Peeters *et al.*, 1988).

Results from studies of cotton fibre nucleoli therefore suggest that the availability of ribosomes, which is substantially controlled by increased rRNA gene transcription and accumulation, is correlated with increased fibre growth. The implication of ribosomal RNA synthesis and accumulation in fibre growth warrants an investigation of rRNA metabolism at the molecular level. There are many levels at which the accumulation of rRNA may be controlled (Timmis *et al.*, 1972; Timmis and Ingle, 1975), including the number of actively transcribed rRNA genes, the rate of transcription, the efficiency of processing the 40S precursor to the mature rRNAs and the stability of the resultant rRNA product.

6.2 ASSESSMENT OF FIBRE LENGTH IN SOME AUSTRALIAN COTTON CULTIVARS

Prior to an investigation of rRNA genes and their transcription in cotton fibres, it was

essential to obtain cotton cultivars which varied significantly in the lengths of their mature fibres. Six varieties were obtained which were thought to differ in their final fibre lengths: *G. hirsutum* cv. Siokra L-22, Siokra 1-4, CS7S, Strumica 5086 and Paymaster 909, and a single *G. barbadense* variety, Pima S-6 (Table 2.1). The Australian commercially cultivated varieties are almost exclusively medium-long staple varieties, of which Siokra 1-4 and Siokra L-22 were the longest at 30.0 and 30.5 mm respectively (CSD Variety Handbook, 1992-1993). At the time of this study, Siokra 1-4 was the most extensively-cultivated commercial variety in Australia. CS7S, with a final fibre length of around 28.5 mm (CSD Variety Handbook, 1992-1993), was the commercial cultivar with the shortest fibres and was developed as a specialist short-season variety for cooler climates. Strumica 5086, an early-maturing variety which originated in Eastern Europe, and Paymaster 909, a short-season variety from Texas, USA, are not cultivated commercially in Australia and are short staple varieties, with fibre lengths of around 23-25 mm (Peter Reid, pers. comm.). Varieties of *G. barbadense*, or Egyptian cotton, are generally classified as long staple and possess fibres with lengths of around 35 mm (Thomson, 1979; Langer and Hill, 1982; Peeters *et al.*, 1987).

Since reliable and accurate fibre length information was only available for the three commercial cultivars, fibre lengths were ascertained in order to confirm previous measurements and verify reputed differences between the fibre lengths. Fibre length for *G. barbadense* cv. Pima S-6 was unable to be determined experimentally because propagation of fruiting plants was unsuccessful under both growth cabinet and greenhouse conditions. Likewise, plants of the cultivar Paymaster 909 failed to thrive and the variety was removed from the analysis, leaving Strumica 5086 as the single short staple cultivar. The method adopted for determination of fibre length in the four remaining *G. hirsutum* varieties was the same as outlined by Schubert *et al.* (1973) and first described by Gipson and Ray (1969). Length measurements were restricted to fibres from the chalazal region of the seed, on the rounded side. Although fibre development within a boll is generally considered to be

synchronous, fibres of the distal (micropylar) half of the seed begin elongation 1-2 days later than those in the chalazal portion (Figure 1.1). Fibre length data for each of four *G. hirsutum* cultivars, Siokra 1-4, Siokra L-22, CS7S and Strumica 5086 are presented in Appendices 1-4, with a summary in Table 6.1.

Table 6.1: Fibre lengths in four cotton varieties

Boll age (DPA)	Cotton variety (*G. hirsutum*) and fibre length in mm			
	Siokra L-22	Siokra 1-4	CS7S	Strumica 5086
5	2.29 ± 0.54	2.28 ± 0.46	2.13 ± 0.34	2.17 ± 0.41
10	9.10 ± 1.79	9.19 ± 1.30	10.17 ± 1.42	8.62 ± 1.25
15	20.31 ± 3.23	20.06 ± 3.42	20.65 ± 2.62	18.32 ± 1.77
20	27.07 ± 2.21	28.96 ± 1.55	24.88 ± 1.27	22.88 ± 1.74
25	29.33 ± 1.79	26.85 ± 1.33	25.05 ± 1.76	21.93 ± 1.14
30	29.35 ± 1.56	28.08 ± 1.07	26.05 ± 1.52	23.30 ± 2.98
35	30.08 ± 1.47	27.85 ± 1.81	26.90 ± 1.61	22.80 ± 1.63
40	29.07 ± 2.02	28.83 ± 1.46	25.94 ± 1.52	22.55 ± 2.17
45	29.55 ± 1.63	28.75 ± 1.22	25.77 ± 1.40	22.87 ± 1.88
50	29.93 ± 1.33	27.75 ± 1.85	25.95 ± 1.99	23.33 ± 2.32

The results, presented graphically in Figure 6.1, indicated that the attained lengths of fibres from each of the three commercial cultivars are in agreement with variety specifications (CSD Variety Handbook, 1992-1993). As expected, seeds of *G. hirsutum* cv. Strumica 5086 had the shortest fibres, with a final length of approximately 23 mm, which differed significantly from the fibre lengths attained in both Siokra L-22 and Siokra 1-4. A significant difference was also obtained between the final lengths of fibres from CS7S and Siokra L-22. All of the four *G. hirsutum* varieties were expected to differ significantly from the long staple fibres of *G. barbadense* Pima S-6. Fibre elongation had virtually ceased by 20 DPA (Figure 6.1) and since the few days prior to the cessation of elongation overlaps with secondary wall synthesis, the results are in agreement with published observations on onset of secondary wall synthesis at approximately 16 DPA. Conformity of this result with those obtained from field-grown plants

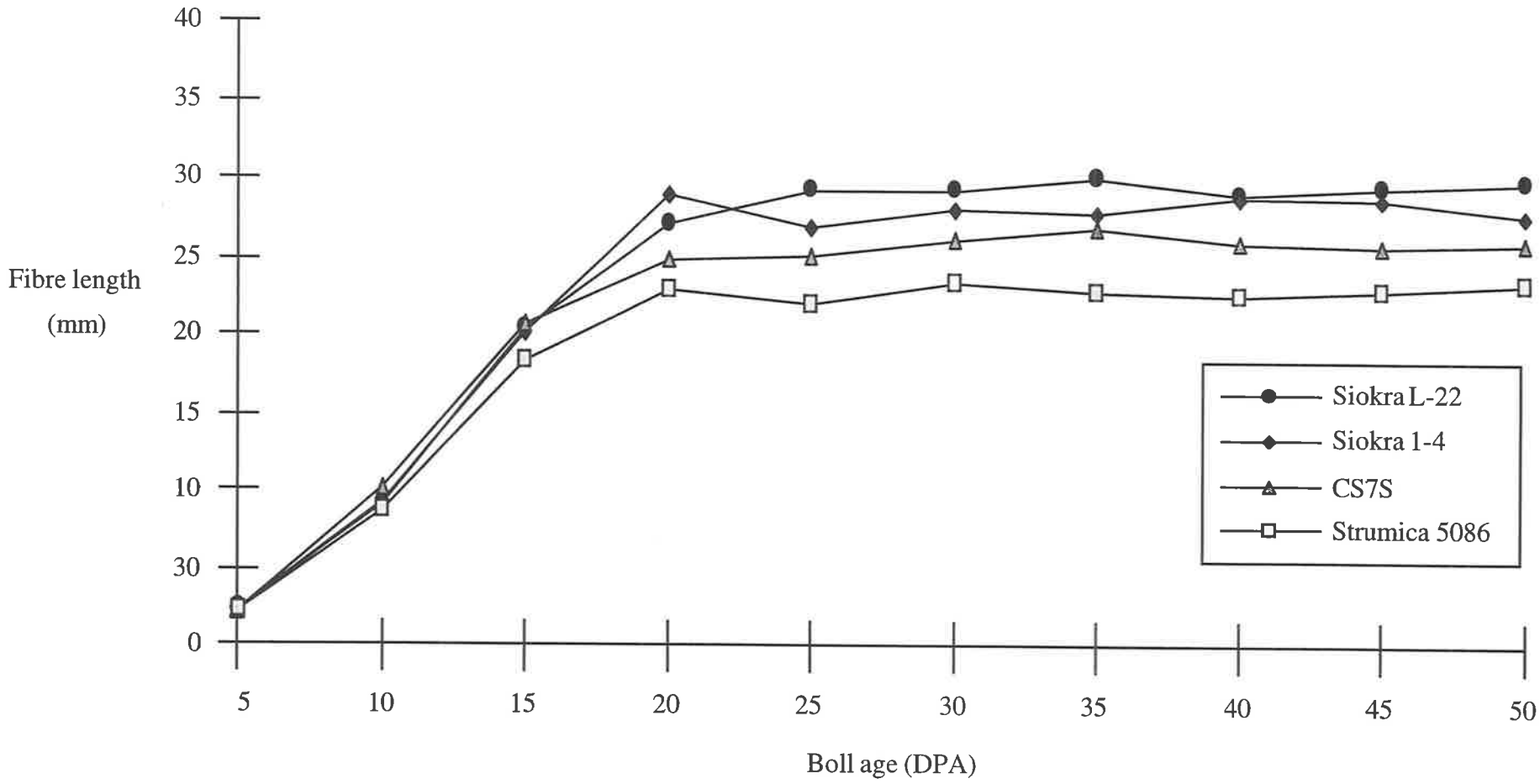


Figure 6.1: Cotton fibre lengths

Graph showing cotton fibre lengths from four varieties, measured from 5 DPA to 50 DPA. Each point on the graph represents an average of fibre measurements from 18-36 different seeds.

suggests that the growth cabinet conditions used in these experiments are compatible with field conditions.

6.3 Analysis of Nucleolar Volume in Young Fibres of Three Cotton Cultivars

The size of the nucleolus has been assumed to reflect its overall activity but may also indicate the degree of coordination between transcription of rRNA genes, processing of primary ribosomes to form nascent ribosomal particles and transport of these particles to nucleoplasm and cytoplasm. Generally, rapid protein synthesis is correlated with an increase in nucleolar size and in some plants, changes in number, size and shape of nucleoli have been noted in differentiating cells (Ghosh, 1987).

During the initial stages of cotton fibre development the nucleus and nucleolus enlarge considerably (Peeters *et al.*, 1987). After reaching a maximum between 2 and 10 DPA, depending on the variety, the size of the nucleolus gradually decreases to attain a constant size at around 20 DPA (Peeters *et al.*, 1987; Peeters *et al.*, 1991). Numerous studies of early fibre development in different cotton cultivars have suggested a close relationship between nucleolar growth and final fibre dimensions. In addition, the observation that both final fibre dimensions and nucleolar volume are normally distributed suggests that large and small nucleoli in a fibre population belong to long and short fibres respectively (DeLange *et al.*, 1978; Peeters *et al.*, 1987; Peeters *et al.*, 1988).

Characterisation of three Australian cotton cultivars was carried out in order to determine whether their fibre development with respect to nucleolar volume is in agreement with that described for other varieties. The nucleolar size in fibres from the *G. hirsutum* varieties Strumica 5086, CS7S and Siokra L-22 was examined at early stages of growth by light microscopy (Figure 6.2). Nucleoli were prominent in fibres aged 2-14 DPA but were difficult to locate after 14 DPA when substantial fibre elongation had occurred. Small nucleoli

Figure 6.2: Nucleoli in developing cotton fibres

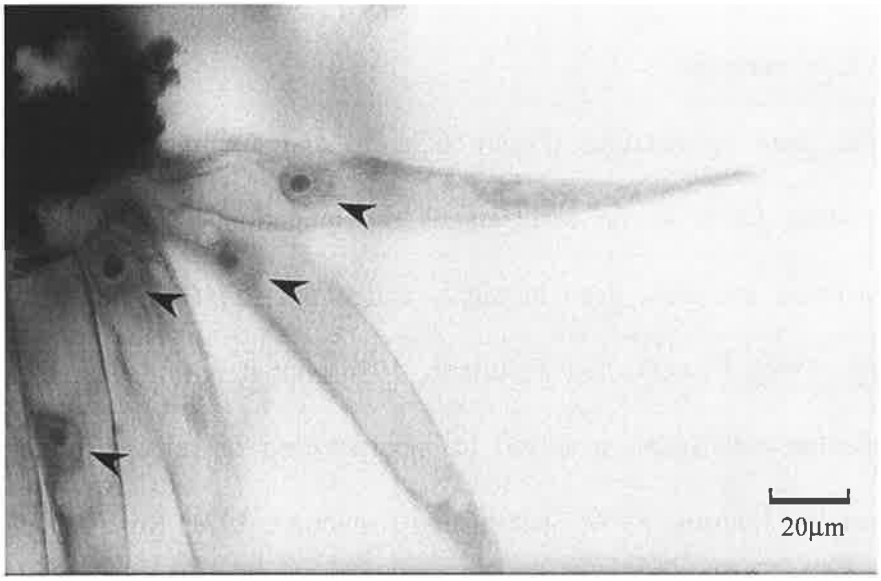
Squashed preparations of stained cotton fibres in the early stages of development, viewed using light microscopy.

a. Fibres aged 2 DPA from *G. hirsutum* cv. Strumica 5086. Extensive cell elongation has occurred and fibres can be clearly seen radiating from the epidermal layer of the ovule wall. The cytoplasm is visible in a thin rim in the fibre tips and between the cell wall and tonoplast, or vacuole membrane, with nuclei and prominent nucleoli (arrowed) located midway down the fibres.

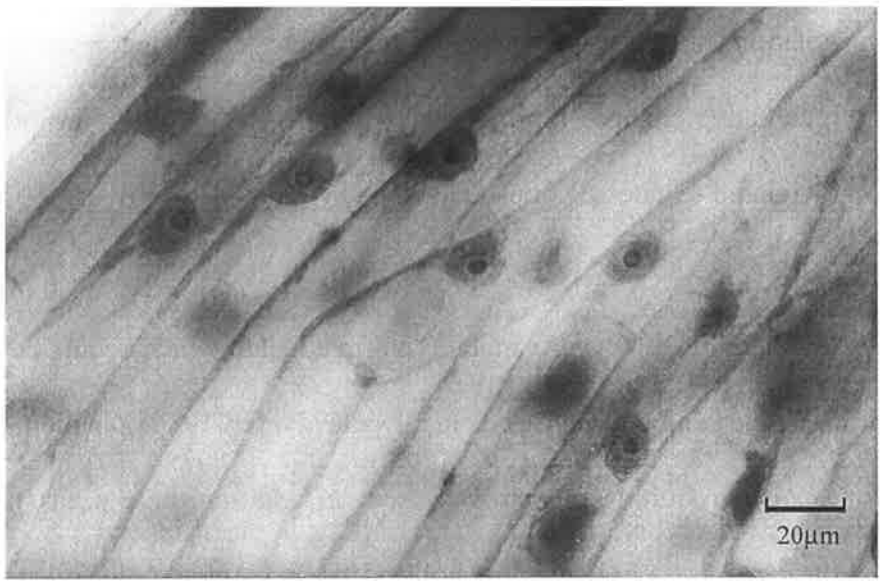
b. Fibres aged 4 DPA from *G. hirsutum* cv. Strumica 5086, showing nuclei and nucleoli, with vacuoles, halfway down the elongated fibre cells.

c. Nucleus and nucleolus from a 10 DPA fibre of *G. hirsutum* cv. CS7S, showing presence of a large, spherical nucleolar vacuole.

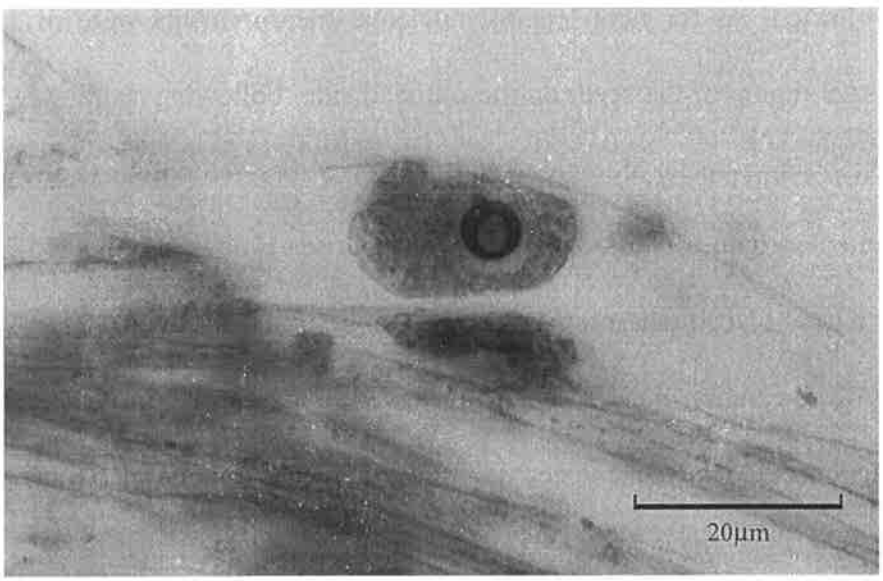
a.



b.



c.



were visible in *G. hirsutum* fibres up to 25-30 DPA but were not observed in 40-50 DPA fibres in any of the three varieties.

Vacuoles were conspicuous (Figure 6.2a, b) and are a common occurrence in the nucleoli of cotton fibres in the early stages of elongation (Ramsey and Berlin, 1976a). Nucleolar vacuoles are often seen in highly active nucleoli (Risueño and Medina, 1986; Risueño *et al.*, 1988; Risueño and Testillano, 1994) but it is not clear from the literature whether nucleolar vacuolation is linked to incorporation or release of nucleolar material Conflicting results (Johnson, 1969; Deltour and Brouchart, 1971; Rose *et al.*, 1972; Shaw *et al.*, 1995) seem to preclude a simple relationship between rRNA gene transcription and the degree of vacuolation of nucleoli. It has been hypothesised that vacuoles result from quantitative or qualitative changes in processing or transport of ribosome precursors rather than being simply related to rates of ribosomal RNA synthesis (Rose *et al.*, 1972; DeLange *et al.*, 1978).

In order to reduce the effects of variability in fibre length which occurs between different bolls on a fruiting branch and between seeds in a locule (Porter, 1936; Davidonis and Hinojosa, 1994), only one boll per fruiting branch was used. In addition, only seeds from a position midway within the carpels of a boll and only bolls of the 2nd, 3rd and 4th fruiting branches of a plant were collected. As for fibre lengths, nucleolar measurements were restricted to fibres from the chalazal region of the seed, on the rounded side. Following staining by methyl green-pyroninY, which stains nuclei blue-green and nucleoli a rose-red colour (Taft, 1951; Kurnick, 1952), the nucleolar diameters of 25 nucleoli from four seeds of each boll aged 2-12 DPA were examined for a total of 100 measurements per sampling. For bolls aged 14 DPA, 25 cells from one seed only were examined and 16 DPA fibres were only inspected for the variety Siokra L-22. The data, presented in Appendices 5-7 and graphically in Figure 6.3, is summarised in Table 6.2.

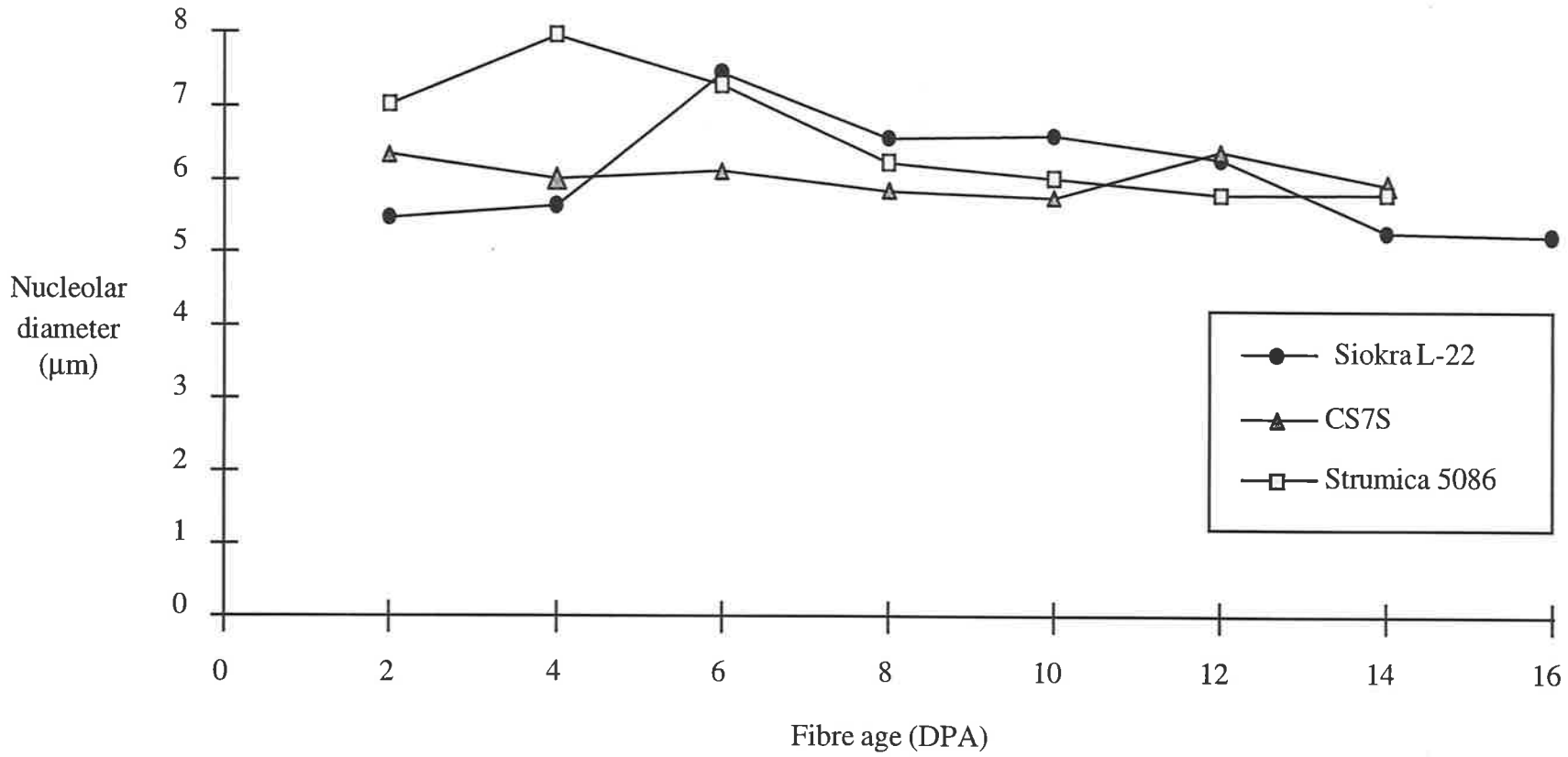


Figure 6.3: Nucleolar growth profiles in three cotton cultivars

Graphical representation of nucleolar growth in young fibres of three cotton cultivars, measured from 2 DPA to 16 DPA at 2 day intervals. Each point on the graph, except for fibres aged 14 and 16 DPA, represents an average of diameter measurements from 100 different nucleoli. Each point for fibres aged 14-16 DPA represents measurements from 25 nucleoli.

Table 6.2: Nucleolar diameter in young fibres of three cotton varieties

Fibre age (DPA)	Cotton variety (*G. hirsutum*) and nucleolar diameter in μm		
	Siokra L-22	CS7S	Strumica 5086
2	5.46 ± 0.60	6.34 ± 0.51	7.02 ± 0.59
4	5.64 ± 0.89	6.01 ± 0.46	7.96 ± 0.75
6	7.46 ± 0.67	6.12 ± 0.75	7.30 ± 0.62
8	6.57 ± 0.52	5.85 ± 0.59	6.24 ± 0.53
10	6.61 ± 0.82	5.76 ± 0.50	6.02 ± 0.58
12	6.30 ± 0.58	6.40 ± 0.67	5.80± 0.55
14	5.29 ± 0.49	5.95 ± 0.37	5.82 ± 0.38
16	5.25 ± 0.61		

The values obtained for nucleolar diameters in this study are consistent with nucleolar volumes obtained in previous studies (Peeters *et al.*, 1991) and several general observations can be made from the results obtained. The nucleolar growth profiles clearly differed between the three *G. hirsutum* varieties Siokra L-22, Siokra 1-4 and CS7S. Nucleolar size in fibres of CS7S appeared to remain constant over the time period examined with a possible peak at approximately 12 DPA (Figure 6.3). The mean nucleolar diameter in Strumica 5086 and Siokra L-22, on the other hand, reached a maximum at approximately 4 DPA and 6 DPA, respectively, before declining (Figure 6.3). This is in agreement with previous studies on varieties of *G. hirsutum*, which have shown that nucleoli increase in size to peak between 2 and 10 DPA, depending on the variety (DeLanghe *et al.*, 1978; Peeters *et al.*, 1987; Peeters *et al.*, 1988).

The largest nucleoli were observed in fibres from Strumica 5086, which has the shortest fibres of the three cultivars (Figure 6.1), and nucleolar size in fibres of Strumica 5086 peaked prior to that of the remaining varieties (Figure 6.3). This is in accord with experiments by Peeters *et al.* (1987) in which nucleoli of a short-staple cotton species, *G. arboreum*, reached a peak early in fibre development. Interestingly, Siokra L-22 and Strumica 5086, which differ in their fibre lengths (Figure 6.1), also differ in their nucleolar volume profiles. The growth of the

nucleolus was delayed in Siokra L-22 when compared with Strumica 5086, and the size of the nucleolus increased more rapidly in Siokra L-22 than in Strumica 5086 (Figure 6.3).

In conclusion, a positive correlation between the maximal nucleolar volume and final fibre length was not apparent. Moreover, particular nucleoli could not be correlated with particular fibres since the changes in nucleolar size are observed within the first 20 DPA after anthesis while fibre development continues until 40-55 DPA, and it is impossible to follow a single fibre in its entire development.

6.4 Estimation of Ribosomal RNA Gene Number in Cotton Cultivars which Differ in Their Final Fibre Length

One possible level at which rRNA gene expression may be controlled is in the number of ribosomal RNA genes per genome. If present, such differences between cotton varieties may cause variation in the ribosome complement between fibre cells which could contribute to differences between their final fibre lengths. The number of rRNA genes in several cotton genomes was investigated by quantitative hybridisation to genomic DNA samples in comparison with known loadings of the cloned rDNA repeat unit. Previous estimates put the number of rRNA cistrons at 300-400 copies per cotton haploid genome (Walbot and Dure, 1976; Baker *et al.*, 1995) but more recent work by Wendel *et al.* (1995) estimated the number of rDNA repeats in diploid and allotetraploid cotton at approximately 3800 rDNA repeats per haploid genome.

6.4.1 Cloning of the rDNA Repeat Unit from Cotton

Use of a heterologous probe in such an experiment would detect the highly conserved 18S, 5.8S and 25S rRNA genes of the rDNA repeat, with the omittance of divergent spacer sequences (Section 1.5.1). A heterologous probe is thus not ideal for a quantitative experiment. In order to detect the entire gene repeat from cotton, a probe specific for the

cotton rRNA gene, including the transcribed and untranscribed spacer regions, was used. An extensive search of the literature failed to provide evidence for the cloning of rRNA genes from cotton other than the 5S rRNA genes (Qi *et al.*, 1988). The rDNA repeat unit from the genome of *G. hirsutum*, cv. Deltapine 90, an improved cultivar developed by Deltapine Australia Pty Ltd (Table 1.1), was therefore cloned.

The structure of the cotton rDNA repeat unit was assessed by hybridising restricted cotton genomic DNA with the rRNA gene from cucumber (Kavanagh and Timmis, 1986). As expected, cross-hybridisation to sequences in the cotton genome was extensive and the probe detected several genomic DNA fragments for each restriction enzyme (Figure 6.4a, lanes 1, 2, 4-10). A smear of hybridisation in the last four tracks is due to partial restriction by enzymes *Pst*I, *Sal*I, *Pvu*II and *Sma*I and in the cases of *Pst*I and *Sma*I, results from the extensive methylation of cytosine residues in plant genomic DNA (Gruenbaum *et al.*, 1981). The exception was *Xba*I, which produced a single main band of hybridisation, sized at approximately 9 kb (Figure 6.4a, lane 3) and it was assumed that the restriction enzyme *Xba*I digests the rDNA cistron once per 9 kb repeat unit. This repeat unit size is in the range reported for other dicotyledonous plants, whose repeat units vary in size from between 8 kb and 12 kb (Vedel and Delseny, 1987) and is compatible with a previous estimate of 9.4 kb for the length of the cotton rDNA repeat unit (Wendel *et al.*, 1995).

Cloning of the size-selected DNA fragment, which was evident as a prominent band on an agarose gel containing cotton genomic DNA restricted with *Xba*I (not shown) yielded two positive clones, pB15 and pB50 (Figure 6.4b), both of which contained inserts of around 9 kb in size (not shown). One of the clones, pB15, was chosen for further analysis and renamed pGhR1. Restriction of DNA from pGhR1 identified several enzymes which contained recognition sites within the insert (Figure 6.5a). At least one DNA fragment in each lane was similar in size to that obtained for the corresponding enzyme on the genomic Southern (Figure 6.4a). For example, the 4 kb *Eco*RI fragment of pGhR1 (Figure 6.5a, lane 11) corresponds to a

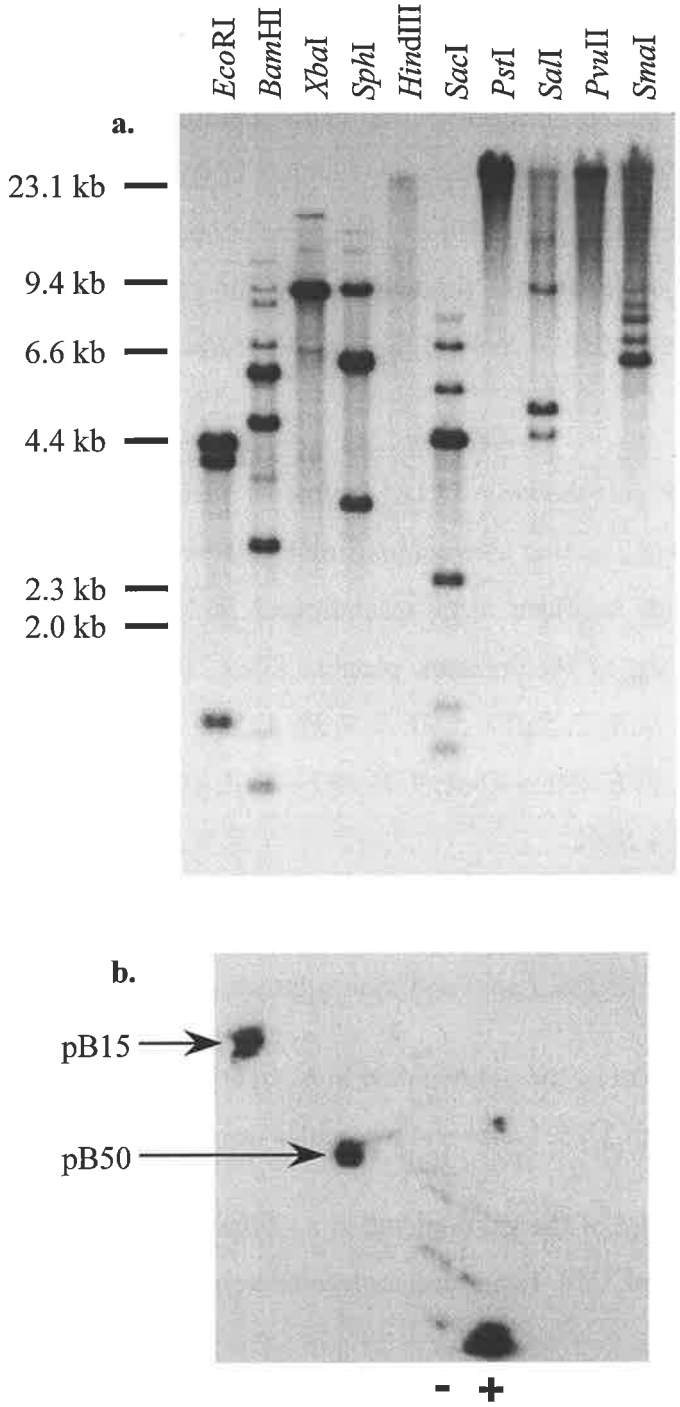


Figure 6.4: Cloning of the rDNA repeat unit from the genome of cotton

a. Genomic Southern blot probed with the insert from cucumber rDNA clone pCU18. Each lane was loaded with 2 µg of total genomic DNA from *G. hirsutum* L. cv. Deltapine 90 restricted with endonucleases as indicated. Numbers on the left represent the molecular weights of wild type Lambda DNA restricted with *Hin*dIII (kb).

b. Grunsteins blot of putative cotton rDNA clones, hybridised with the pCU18 insert. Positive clones pB15 and pB50 are indicated by arrows, and the positions of positive and negative control colonies are designated by a (+) and (-) respectively.

Figure 6.5: Physical mapping of the cotton rDNA repeat unit

a. Agarose gel showing DNA fragments resulting from restriction of pGhR1. The first lane contains molecular weight markers (Lambda DNA restricted with *Hin*dIII) with fragment sizes as indicated in kb. The remaining lanes each contain approximately 2 μg of recombinant plasmid DNA digested with the following restriction endonucleases: 1. *Xba*I; 2. *Sph*I; 3. *Sac*I; 4. *Sma*I; 5. *Sal*I; 6. *Pvu*II; 7. *Hin*dIII; 8. *Eco*RV; 9. *Kpn*I; 10. *Bam*HI; 11. *Eco*RI; 12. *Cla*I and 13. *Pst*I.

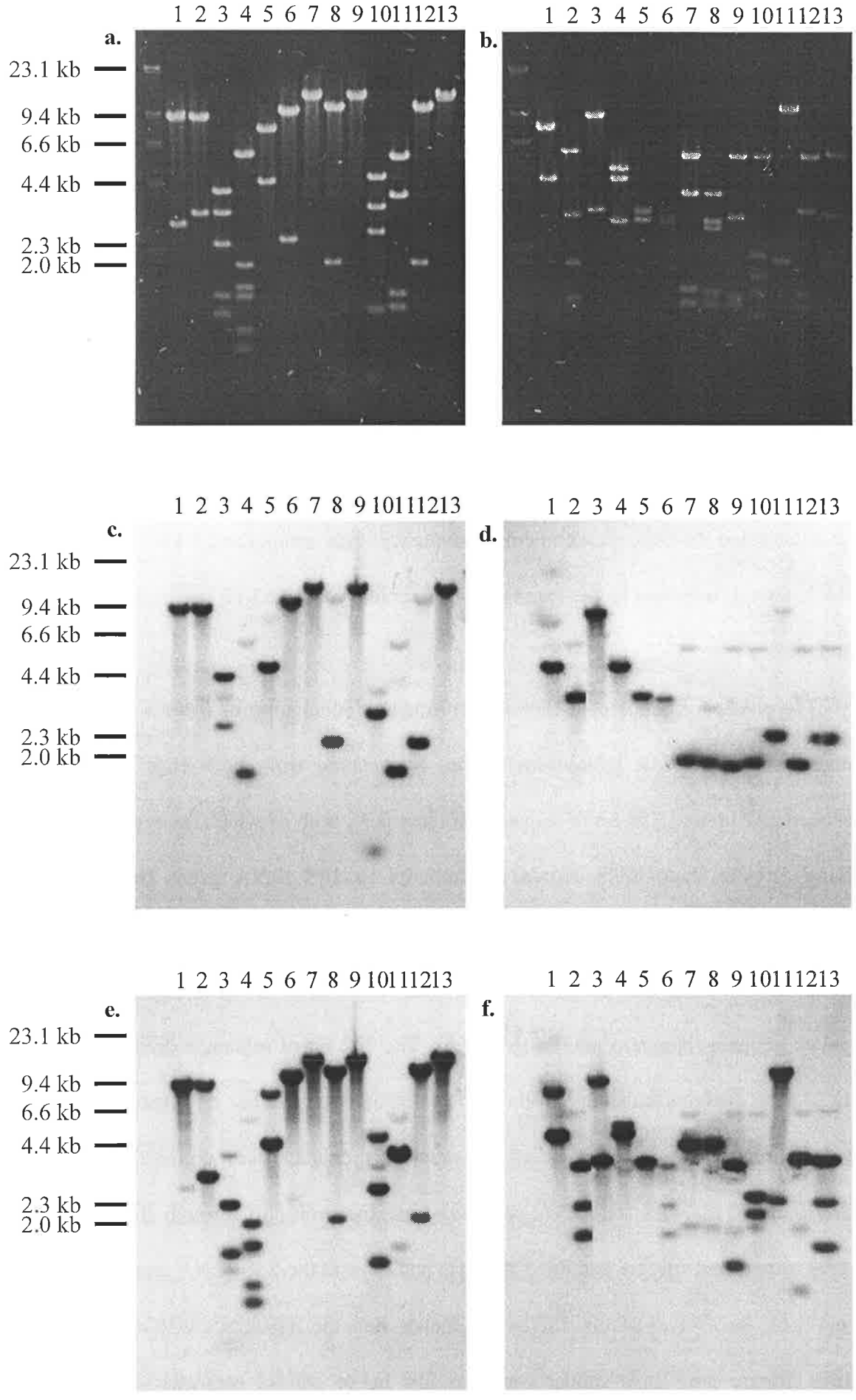
b. Agarose gel showing DNA fragments resulting from restriction of pGhR1 for restriction mapping. The first lane contains molecular weight markers (Lambda DNA restricted with *Hin*dIII) with fragment sizes as indicated in kb. The remaining lanes each contain approximately 2 μg of recombinant plasmid DNA, digested with the following restriction endonucleases: 1. *Sal*I; 2. *Sal*I + *Sph*I; 3. *Sph*I; 4. *Sal*I + *Xba*I; 5. *Sph*I + *Xba*I; 6. *Sal*I + *Xba*I + *Sph*I; 7. *Eco*RI; 8. *Eco*RI + *Xba*I; 9. *Eco*RI + *Sal*I; 10. *Eco*RI + *Sph*I; 11. *Cla*I; 12. *Eco*RI + *Cla*I and 13. *Cla*I + *Sph*I.

c. Southern blot of the gel depicted in a., hybridised with the 18S cucumber rRNA gene derived from clone pCU18. Lanes and molecular weight markers are labelled as above.

d. Southern blot of the gel depicted in b., hybridised with the 18S cucumber rRNA gene derived from clone pCU18. Lanes and molecular weight markers are labelled as above.

e. Southern blot of the gel depicted in a., hybridised with the 25S cucumber rRNA gene derived from clone pCU18. Lanes and molecular weight markers are labelled as above.

f. Southern blot of the gel depicted in b., hybridised with the 25S cucumber rRNA gene derived from clone pCU18. Lanes and molecular weight markers are labelled as above.


1 2 3 4 5 6 7 8 9 10 11 12 13
a.
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
b.
c.
d.
e.
f.

genomic DNA fragment of a similar size (Figure 6.4a; lane 1) which hybridised strongly to the heterologous rDNA probe.

The restriction enzymes which contained three or fewer sites in the cotton rDNA repeat, namely *Eco*RI, *Sph*I, *Sal*I and *Cla*I, were used to map clone pGhR1 (Figure 6.5b; Figure 6.6) and the clone was physically mapped using probes specific for each gene region, derived from the cucumber rDNA repeat clone pCU18 (Figure 6.5c, d, e, f). The insert size of pGhR1 derived from restriction mapping was estimated at 9.8 kb. Probing with the 25S rRNA gene located the cotton 25S sequence to the 3.8 kb EcoRI fragment of pGhR1 (Figure 6.5e, f). The 25S gene terminates within the 0.9 kb *Eco*RI/*Sal*I fragment and the 0.6 kb *Cla*I/*Eco*RI fragment of pGhR1 (Figure 6.6). Probing with a fragment of pCU18 specific for the 18S rRNA gene determined the corresponding cotton sequence to be within the 1.5 kb *Eco*RI fragment of pGhR1, directly adjacent to the fragment which contains the 25S rRNA gene (Figure 6.5c, d; Figure 6.6).

The physical map of pGhR1 was confirmed by subcloning of the 1.5 kb EcoRI which contains the 18S rRNA gene from cotton. Sequencing from both ends of the subclone produced 382 bp and 256 bp of sequence (Figure 6.7), both of which, as expected, showed a striking (greater than 90%) degree of similarity to 18S rRNA genes from many plants (GenBank and EMBL; September 1994). A more recent search of the databases (GenBank and EMBL; March 1996) identified a 1757 bp 18S rRNA gene recently sequenced from a *G. hirsutum* genome (Nickrent and Soltis, 1995). The 382 bp of sequence derived from pGhR1 showed 98% nucleotide identity to the cotton 18S rRNA gene but terminated approximately 250 bp short of the 3' terminus of the previously characterised cotton 18S gene. Since transcription of the 40S pre-rRNA proceeds from the promoter through the 18S gene, the internal transcribed spacers and the 5.8S gene and beyond the 25S rRNA gene (right to left in Figure 6.6), the 3' end of the 18S gene extends into the adjacent *Cla*I/*Eco*RI fragment of pGhR1 (Figure 6.6). In a similar way, the 256 bp of pGhR1 sequence derived from the

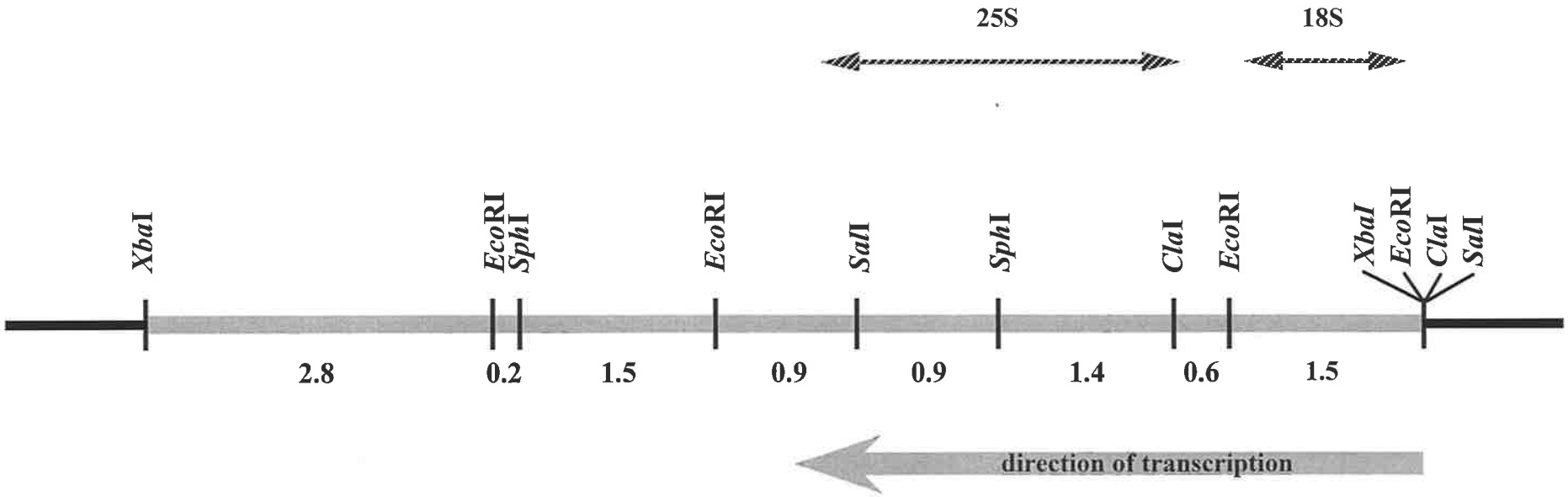


Figure 6.6: Partial restriction map of the cotton rDNA repeat clone pGhR1

Restriction sites of the enzymes *Eco*RI, *Sph*I, *Sal*I and *Cla*I are indicated, with fragment sizes given in kb and the regions hybridised by probes specific for the 25S and 18S rRNA genes designated by arrows.

a.

```
    TGCAACAAACCCCGACTTCTGGAAGGGATGCATTTATTAGATAAAAGGTCGACGCGGCTT
  1 ---------+---------+---------+---------+---------+---------+ 60

    TGCTGCTCTGATGATTCATGATAACTCGACGGATGCGACGGCCTTTGTGCGGCGACGCAT
 61 ---------+---------+---------+---------+---------+---------+ 120

    CATTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCTACTATGGTGGTGA
121 ---------+---------+---------+---------+---------+---------+ 180

    CGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGCTACCACATC
181 ---------+---------+---------+---------+---------+---------+ 240

    CAAGGAAGGCAGCAGG
241 ---------+------ 256
```

b.

```
    CTCGTTGAAGACCAACAATTGCAATGATCTATCCCCATCACGATGAAATTTCAAAGATTA
  1 ---------+---------+---------+---------+---------+---------+ 60

    CCCGGGCCTGTCGGCAAGGCTATAGACTCGTTGAATACATCAGTGTAGCGCGCGTGCGCC
 61 ---------+---------+---------+---------+---------+---------+ 120

    CAGAACATCTAAGGGCATCACAGACCTGTTATTGCCTCAAACTTCCTTGCCCTAAAAGGC
121 ---------+---------+---------+---------+---------+---------+ 180

    ATAGTCCCTCTAAGAAGCTAGCCACGGAGGATCACCTCCGTGTAGCTAGTTAGCAGGCTG
181 ---------+---------+---------+---------+---------+---------+ 240

    AGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCGCTCCACCAACTAAGAACAGGCCA
241 ---------+---------+---------+---------+---------+---------+ 300

    TGCACCACCACCCATAGAATCAAGAAAGAGCTCTCAGTCTGTCAATCCTTACTATGTCTG
301 ---------+---------+---------+---------+---------+---------+ 360

    GACCTGGTAAGGTTCC
361 ---------+------ 376
```

Figure 6.7: Sequence of the 18S gene region of cotton rDNA repeat clone pGhR1

a. Partial sequence of the 18S subclone of pGhR1, from the (-40) primer of pBluescript®SK(-).

b. Partial sequence of the 18S subclone of pGhR1, from the RSP of pBluescript®SK(-). The reverse complement of the sequence depicted above showed homology to 18S rRNA sequences.

opposite end of the subclone (Figure 6.7b) showed 98% nucleotide identity to the 5' end of the cotton 18S rRNA gene (Nickrent and Soltis, 1995), but terminated approximately 170 bp short of the 5' end of the gene. Since *Xba*I defines a single rDNA repeat unit, the remaining 170 bp of the 18S gene should be located at the other end of the pGhR1 clone, contained in the 2.8 kb *Xba*I/*Eco*RI fragment (Figure 6.6). The small amount of hybridisation which occurred to the 2.8 kb *Xba*I/*Eco*RI fragment by the 18S probe (Figure 6.5d, lane 8) lends support to this hypothesis, but the lack of similar hybridisation to the 3.0 kb *Xba*I/*Sph*I (Figure 6.5d, lane 5) and 5.4 kb *Xba*I/*Sal*I (Figure 6.5d, lane 4) fragments of pGhR1 suggest that the remaining portion of the 18S gene may be present on a small *Xba*I fragment which was not detected by Southern analysis (Figure 6.4a, lane 3). The lengths of ITS1 and ITS2, which flank the 5.8S coding region (Figure 1.4) are approximately 180 bp to 250 bp in plants, including cotton (Torres *et al.*, 1990; Wendel *et al.*, 1995) and are compatible with the physical map of pGhR1 in Figure 6.6.

The sequence differences observed between the cotton 18S genes may be due to sequencing errors, since only one DNA strand of pGhR1 was sequenced, or may reflect intraspecific sequence heterogeneity in the 18S rRNA genes of cotton. Examples of polymorphism in the rDNA coding regions of other organisms have been reported (Kolosha *et al.*, 1986), but to the contrary, the 5.8S gene shows strong sequence conservation in several diploid and tetraploid cotton varieties (Wendel *et al.*, 1995), making the former explanation most likely.

6.4.2 The Diphenylamine Reaction for Quantitation of DNA

A quantitative hybridisation experiment is reliant on precise determination of DNA amounts in solution. UV spectrophotometry at a wavelength of 260 ηm is a standard procedure for determination of DNA concentration which is dependent upon the nature and quantity of other components in the extract. Since both the cotton genomic DNA samples and

the plasmid clone pGhR1 were known to be contaminated with RNA, the Burton (1956) diphenylamine reaction was employed to accurately determine the concentration of cotton genomic and pGhR1 clone DNA against a known standard. The diphenylamine reaction is a simple spectrophotometric method which is widely utilised for estimating DNA content in tissue extracts and although not entirely specific for deoxyribose, the reaction is not affected by carbohydrates present in the extract (Ingle, 1963).

Pilot reactions were performed to calculate the range of standards required such that the values obtained for the unknowns, adjusted to a concentration of approximately 1 μg/μl using the reading from UV spectrophotometry, fell within the range of the standards. The DNAs of unknown concentration (cotton genomic DNA and pGhR1 plasmid DNA) were then run in triplicate against the known standards, which were also run in triplicate. Final data from the diphenylamine reaction is presented in Appendix 8a. The regression calculated for the standard curve was used to derive the unknown DNA concentrations in μg of DNA per μl (Appendix 8b; column 7). DNA concentration calculated by the diphenylamine reaction was invariably less than half the value calculated by standard UV spectrophotometry, most likely due to RNA contamination of the DNA samples which would elevate the spectrophotometric reading at 260 ηm. Following determination of DNA concentration using the diphenylamine reaction, DNA solutions were diluted and adjusted to standard concentrations to be used throughout the hybridisation experiments.

6.4.3 Assessment of Cross-hybridisation by pGhR1 to Plastid rDNA Sequences

Higher plants are characterised by three distinct genomes located in different cellular compartments: the nucleus, chloroplasts and mitochondria. Ribosomes exist in the organelles in addition to the cytoplasm, and the constituent rRNAs are encoded by the organellar genomes. The plastid (chloroplast) genome is represented by a circular DNA molecule of approximately 150 kDa (Palmer, 1985; Shinozaki *et al.*, 1986), within which are two inverted repeats of a 20

kb sequence containing the plastid-specific ribosomal RNA genes. The number of chloroplasts per cell and the number of genomes per chloroplast varies widely between cells and is largely dependent upon the age of the tissue (Scott and Possingham, 1980). Quantitation of the number of ribosomal RNA genes in a particular nuclear genome therefore requires a probe specific for the nuclear rRNA genes, to the exclusion of the plastid-derived homologues. Evidence suggests that the distinctive rRNA species encoded by the mitochondrial genome are divergent from their nuclear counterparts (Gray *et al.*, 1983) and were therefore not expected to hybridise the nuclear-derived pGhR1. pGhR1 did not hybridise mitochondrially-derived rRNA transcripts isolated from a cotton fibre cDNA library (Section 3.3).

An experiment designed to assess the cross-hybridisation of pGhR1 to plastid rDNA sequences exploited the differences in methylation status between the nuclear and organellar genomes. The nuclear genome of many higher plants is extensively methylated at cytosine residues (Doerfler, 1983; Gruenbaum *et al.*, 1981) such that 5-methylcytosine may account for 25% to 35% of cytosine residues. In contrast, the chloroplast DNA of these species is generally unmethylated (Smillie and Scott, 1969; Scott and Possingham, 1980). Use of the restriction endonuclease *Hpa*II, which cannot cleave CCGG sequences when the internal cytosine residue is methylated, distinguishes between the two methylation states (Nelson and McClelland, 1987). Restriction of total plant DNA with *Hpa*II results in cleavage of the unmethylated plastid genome while the extensively methylated nuclear genome remains essentially intact. Total DNA from the six experimental cotton cultivars, *G. hirsutum*, cv. Siokra 1-4, Siokra L-22, CS7S, Strumica 5086, Paymaster 909 and *G. barbadense*, cv. Pima S-6 in addition to *G. hirsutum* cv. Deltapine 90 was restricted with *Hpa*II and *Xba*I, and the resultant Southern probed with the insert from pGhR1 (Figure 6.8a, b). Identical hybridisation patterns were produced in the six *G. hirsutum* varieties (Figure 6.8b, lanes 1-6) and in the single *G. barbadense* variety (Figure 6.8b, lane 7), and as expected, cotton rDNA repeat units were cleaved inefficiently by *Hpa*II but were digested to completion with *Xba*I. Most of the

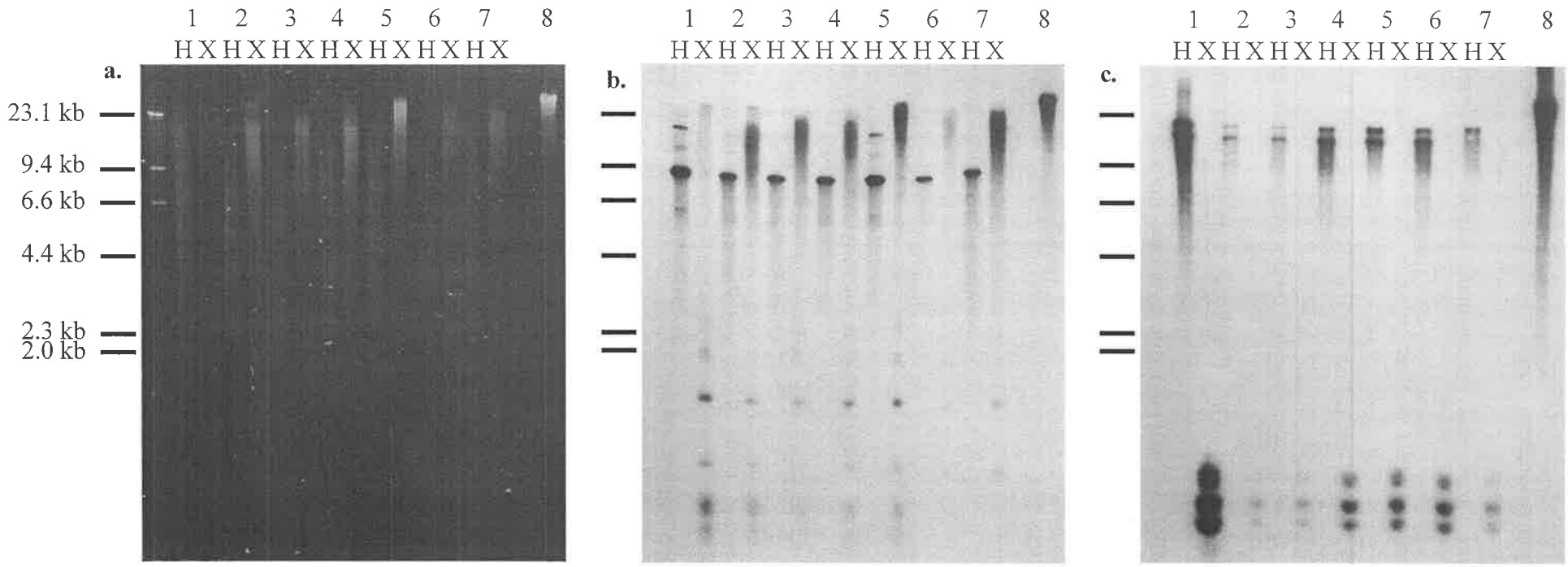


Figure 6.8: Assessment of cross-hybridisation to plastid rDNA sequences

a. Agarose gel electrophoresis of total DNA from several cotton genomes. Lanes labelled X and H contain 1 µg of DNA restricted with *Xba*I and 2 µg of DNA restricted with *Hpa*II respectively and the cotton varieties are: 1. Deltapine 90; 2. Strumica 5086; 3. Paymaster 909; 4. CS7S; 5. Siokra 1-4; 6. Siokra L-22 and 7. Pima S-6. The first lane contains molecular weight markers (Lambda DNA restricted with *Hin*dIII) with fragment sizes as indicated in kb and lane 8 contains 2 µg of unrestricted DNA from *G. hirsutum* cv. Deltapine 90.

b. Southern blot of the gel depicted in a., hybridised with the 9 kb insert from pGhR1. Lanes and molecular weight markers are labelled as above.

c. Southern blot of the gel depicted in a., hybridised with the 3.1 kb insert from spinach chloroplast rDNA clone Xho3. Lanes and molecular weight markers are labelled as above.

hybridisation occurred to the 9 kb repeat unit band in the *Xba*I-restricted DNA (Figure 6.8b, lane X). Hybridisation of pGhR1 to *Hpa*II-restricted DNA occurred mainly to high molecular weight, unrestricted DNA (Figure 6.8b, lane H) which presumably represents rRNA genes derived from the nuclear genome. In addition to the high molecular weight smear of DNA, three distinct low molecular weight bands are evident, which were absent from the control of unrestricted genomic DNA (Figure 6.8b, lane 8). The bands may be attributed to unmethylated nuclear DNA or to rRNA genes derived from the plastid genome.

Evidence for the former explanation came from Southern hybridisation using an rDNA probe derived from the spinach chloroplast, Xho3 (Crouse *et al.*, 1978; Palmer and Thompson, 1981). Probing of *Hpa*II-restricted cotton genomic DNAs with Xho3 produced three distinct low molecular weight bands of hybridisation (Figure 6.8c, lane H) which are plastid in origin. The bands did not correspond in size to those observed when *Hpa*II-cut genomic DNA was probed with nuclear rDNA sequences (Figure 6.8b, lane H). The low molecular weight fragments which resulted from hybridisation with pGhR1 (Figure 6.8b) conceivably represent a proportion of the rRNA genes which are nuclear in origin but are unmethylated. Methylation of plant nuclear rRNA genes at CpG, CpCpG or CpNpG motifs is widely considered to play a role in the regulation of their transcription (Scott *et al.*, 1984; Section 1.5.4). In plants with a relatively large number of rDNA repeats, a high percentage of the rDNA (greater than 70%) appears completely methylated at these sites and is probably packaged in a highly condensed and transcriptionally inactive heterochromatic form (Hemleben and Zentgraf, 1994). The small percentage of repeats which is completely unmethylated are proposed to represent the actively transcribed portion of rDNA.

In addition, only after extended exposure was a small amount of hybridisation observed to unrestricted nuclear DNA by the plastid probe (not shown), and no hybridisation was observed to the 9 kb *Xba*I fragment which contains the rDNA repeat unit from the cotton nucleus (Figure 6.8c, lane X). Hybridisation of Xho3 occurred to high molecular weight bands

in the *Xba*I restrictions (Figure 6.8c, lane X) which correspond to rRNA genes located on the plastid genome. Interestingly, variation between the chloroplast rRNA genes of the *G. hirsutum* varieties and those of Pima S-6 was evident in that Pima S-6 DNA lacks the second band of hybridisation to Xho3 (Figure 6.8c, lane X). It may be that the chloroplast rDNA repeat units in cotton have an *Xba*I polymorphism.

In summary, the implications from this experiment are that chloroplast rDNA sequences do not cross-hybridise rRNA genes of nuclear origin and, likewise, pGhR1 does not cross-hybridise chloroplast rRNA genes. Quantitation of rRNA genes in total cotton DNA utilising pGhR1 should not therefore be confused by cross-hybridisation to chloroplast sequences.

6.4.4 Assessment of Cross-hybridisation to Contaminant RNA and Vector Sequences

The slotblotting procedure involves treatment of DNA samples with NaOH, which should hydrolyse contaminant RNA in the genomic and plasmid DNA samples. A pilot experiment was undertaken to determine whether this was indeed the case and to ascertain the effects of RNA contamination on the quantitation experiment. Two 4 μg samples of total RNA from leaf tissue were applied to a slotblot, one of which was denatured with formaldehyde in the usual way for RNA slotblotting whilst the other underwent treatment which was identical to that used for the DNA samples. The rRNA transcripts in the denatured RNA sample exhibited strong hybridisation to the pGhR1 probe (not shown) whereas the latter sample did not produce a hybridisation signal (not shown). It appears that the RNA in the treated sample did not adhere to the nitrocellulose membrane and was probably hydrolysed by the NaOH used in preparation of the sample for slotblotting. In addition, Southern analysis showed that the purified pGhR1 insert did not cross-hybridise to vector (pBluescript®SK(-)) DNA, even with an overload of pBluescript®SK(-) DNA and with extended exposure to X-ray film (not shown).

In summary, the control reactions demonstrated that hybridisation by pGhR1 to slotblot

DNA samples can be fully attributed to rRNA genes present in the nuclear genome of cotton. Hybridisation to rRNA genes of organellar genomes or to rRNA gene transcripts present in the contaminating RNA can be assumed to be negligible.

6.4.5 Estimation of rRNA Gene Number in Several Cotton Cultivars

The number of rRNA genes in several cotton genomes was investigated by quantitative hybridisation to genomic DNA samples in comparison with known loadings of the cloned rDNA repeat unit. A calibration series of pGhR1 was devised which would allow regression analysis of a standard curve and subsequent calculation of the number of rRNA genes in each DNA sample, based on intensity of hybridisation to the pGhR1 ribosomal probe. The experiment utilised a slotblotting apparatus, onto which were loaded duplicate 1 µg samples of each of the genomic DNAs. Calculation of the number of cotton genomes in 1 µg of DNA is in Appendix 9 and was estimated assuming 3.8 pg of DNA per cotton tetraploid genome (Geever, 1980).

It is noteworthy that estimates of the genome size of cotton by several different workers using different techniques vary widely. For example, Walbot and Dure (1976) estimated the DNA amount at 0.795 pg of DNA per haploid cell, while Geever's (1980) estimate was 0.95 pg of DNA per haploid cell. In a more recent study, Galau *et al.* (1988) revaluated the data of the previous two workers to arrive at a value of 0.94-1.0 pg of DNA per haploid genome, a result which was also obtained by Firoozabady (1986). Michaelson *et al.* (1991) estimated the haploid genome size of *G. hirsutum* cv. Tamcot CAMD-E at around 3.03 pg of DNA and 2.8 pg of DNA, using scanning microspectrophotometry and flow cytometry respectively. Both these values clearly differ from those of the previous workers. Other estimates, also determined by flow cytometry, vary from 2.2 to 3 pg of DNA per 1*C* DNA amount, depending on the particular variety under study and the standard DNA used (Bennett and Leitch, 1995). Using Feulgen cytophotometric techniques, Edwards *et al.* (1974) and Kadir (1976) obtained 5-fold

differences between the genome sizes of *G. hirsutum* and *G. barbadense*. However, variation in genome size may simply reflect technical shortcomings or may be due to certain assumptions made in computing genome size and are to be viewed with caution (Bennett and Leitch, 1995). For the purposes of this experiment, Geever's (1980) results were used since they have been confirmed independently by Galau *et al.* (1988).

As for the diphenylamine reaction, a series of pilot experiments was carried out in order to calculate the range (DNA amounts) required of the standards. The resultant films were read using an LKB UltroScan XL Laser Densitometer to produce a series of peaks, each of which corresponded to a slot of DNA on the filter (not shown). The measure of absorbence is generally the integrated area under each peak, but in this case appeared to be an unsuitable measure because the baseline was not appropriate for all peaks. In order to ascertain a suitable method for measuring the densitometer output, 24 peaks were selected at random and the peak height correlated with both the weight of the paper in the cutout peak (which served as an accurate measure of the area under the peak and was calculated from the base of the peak) (Figure 6.9a), and the integrated area, as calculated from the baseline set by the LKB densitometer software (Figure 6.9b). The Pearson Correlation Coefficients r_1 and r_2 for weight and integrated area respectively (Bailey, 1959; Figure 6.9) were both statistically different from 0 and from each other, yet only r_2 was statistically different from 1. In conclusion, integrated area and area under the peak are different. Poor correlation between peak height and integrated area suggests that the latter should not be used as a measure of peak area, as it is an artifactual value generated by the immovable LKB scanning software. In contrast, there is good correlation between peak height and peak area, and peak height can therefore be used to estimate the area under the peak.

Two experimental slotblots were loaded independently, each containing two dilution series of plasmid pGhR1, both in duplicate. Each filter contained 1 µg samples of cotton genomic DNAs, also in duplicate and positioned at random on the slotblot. Following

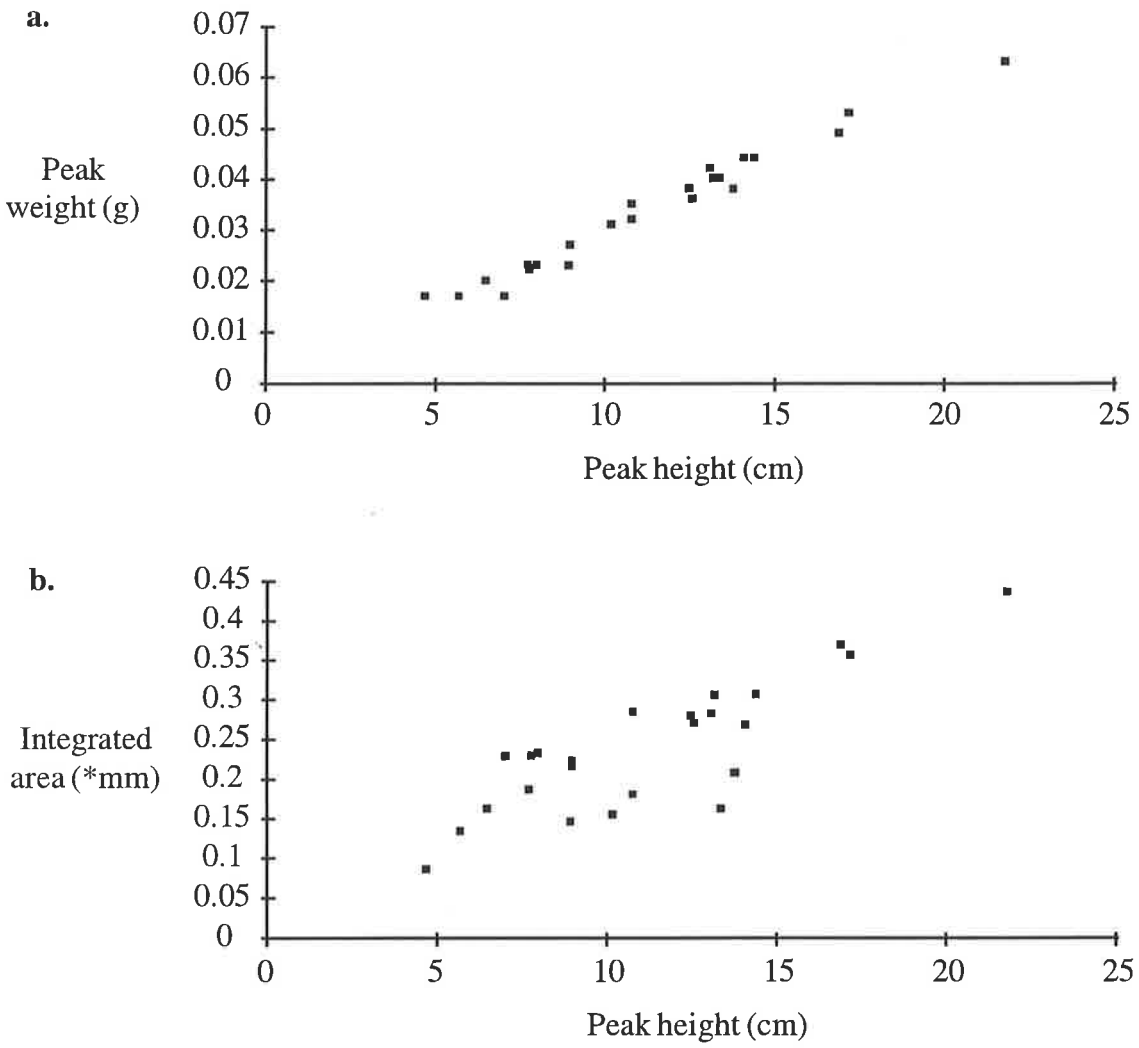


Figure 6.9: Interpretation of the output from scanning densitometry for rRNA gene estimation

a. Correlation between peak height and peak area (as measured by the weight of each peak) of 24 randomly-selected densitometer peaks. The calculated Pearson Correlation Coefficient (r_1), a measure of the linear association between the two variates, is 0.98875, on 22 degrees of freedom (Bailey, 1979).

b. Correlation between peak height and integrated area (as calculated by LKB software) of 24 randomly-selected densitometer peaks. The calculated Pearson Correlation Coefficient (r_2) for this set of data is 0.88025, on 22 degrees of freedom (Bailey, 1979).

hybridisation by the insert from pGhR1 (not shown), an appropriate exposure to X-ray film was scanned twice using the densitometer. The data for filters 1 and 2, which includes standard values and peak heights, are presented in Appendices 10 and 11 respectively. Two sets of data were removed from the analysis of each cotton variety on each filter. Samples for which the hybridisation signal was diffuse or uneven across the slot were omitted, or the slot with the lowest value was removed, since in these cases the DNA may have been inefficiently vacuumed onto the slot. Statistical analysis of the peak height measurements was carried out by the Statistical Consulting Group, University of Adelaide, using repeated measures analysis of variance (Appendix 12).

There are several conclusions which can be drawn from statistical analysis of the data. Firstly, there are significant differences between the two sets of densitometry scans, represented by data set #1 and data set #2 (data set #2 is generally higher than data set #1; Appendix 10, 11) and the four sets of results were separated (Appendix 13). Consequently, final calculations of gene numbers (Table 6.3) were based on the output from statistical analysis (boxed in Appendix 12) using four independent calibration curves (Appendix 13).

Table 6.3: Estimation of rRNA gene copy number in the cotton genome

Cotton variety	Data set and rRNA gene number per unreplicated tetraploid nucleus				Summary		
	Filter #1, data set #1	Filter #1, data set #2	Filter #2, data set #1	Filter #2, data set #2	Average gene number	St. dev.	Fibre length (mm)
Paymaster 909	7647	7886	6592	6705	7208	654	23
Strumica 5086	7592	7832	6531	6639	7149	659	23
CS7S	5772	6062	4497	4459	5198	839	26
Siokra 1-4	6782	7044	5625	5668	6280	739	28
Siokra L-22	6355	6629	5148	5157	5822	781	30
Pima S-6	8555	8769	7606	7792	8181	568	33

Notably, the difference between data sets did not cause any significant differences between the results obtained for each variety nor to their differences in gene number relative to

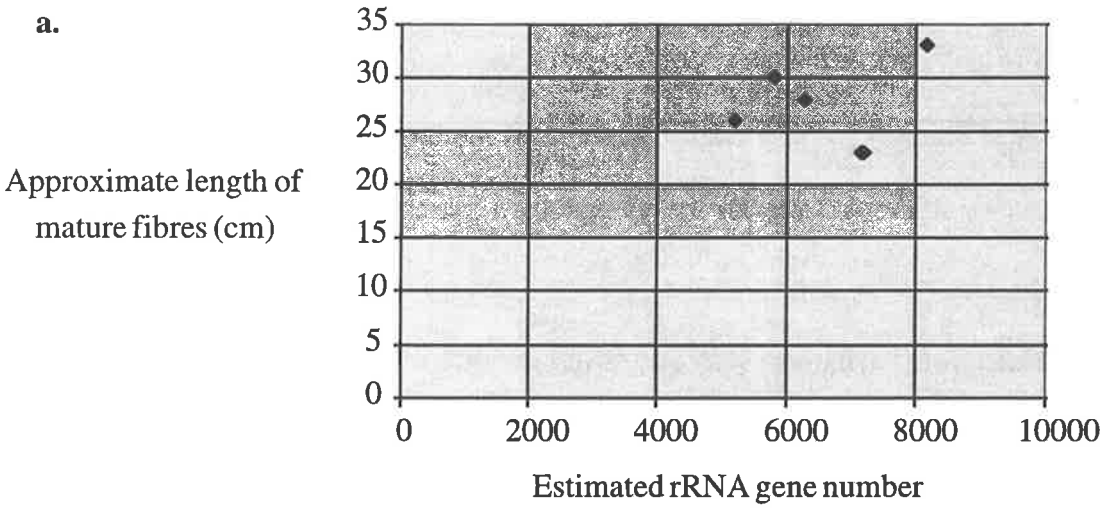
each other (Table 6.3). There are no significant differences in rRNA gene number between plants of the same variety as has been reported in several plant species (Rogers and Bendich, 1987a, Rogers and Bendich, 1987b).

Significant differences were observed between the cotton varieties in rRNA gene copy number. The genome of Pima S-6 contains the highest number of rRNA genes (Table 6.3) and differs significantly from all the *G. hirsutum* varieties except Paymaster 909 and Strumica 5086. The rRNA gene number estimated for CS7S also differs significantly from values obtained for Paymaster 909 and Strumica 5086. The remaining varieties do not differ significantly from each other and in the case of Paymaster 909 and Strumica 5086, are virtually identical. The two varieties which differ significantly in their final fibre lengths (Figure 6.1), Siokra L-22 and Strumica 5086, do not have different rRNA gene numbers. Intraspecific variation in rDNA number in *G. hirsutum* is not surprising and is common in plants. For example, the number of rRNA genes has been found to vary by as much as 95-fold between individuals within a single population of broad bean, *Vicia faba* (Rogers and Bendich, 1987b). The numbers of rRNA genes do not, however, correlate with final fibre length of the cotton varieties under study (Figure 6.10a).

There are significant differences between the results on each filter in terms of absolute gene numbers (Table 6.3). The values obtained from Filter #2 are generally lower and are not corrected by the standards. Standard values between the two filters differ by approximately 0.3, whereas the genomic DNA samples differ in their absorption by approximately 2. This is most likely caused by pipetting or dilution errors for the calibrations on one of the filters and indicates that absolute gene number remains an unknown quantity. In a relative sense however, the filters give the same answer, best visualised as a correlation (Figure 6.10b) and statistically, the filters are consistent with respect to differences in gene numbers between the varieties.

It should be noted that the calculation of rDNA copy number includes the assumptions that the *G. hirsutum* and *G. barbadense* cultivars have the same genomic size as those in cited

a.



b.

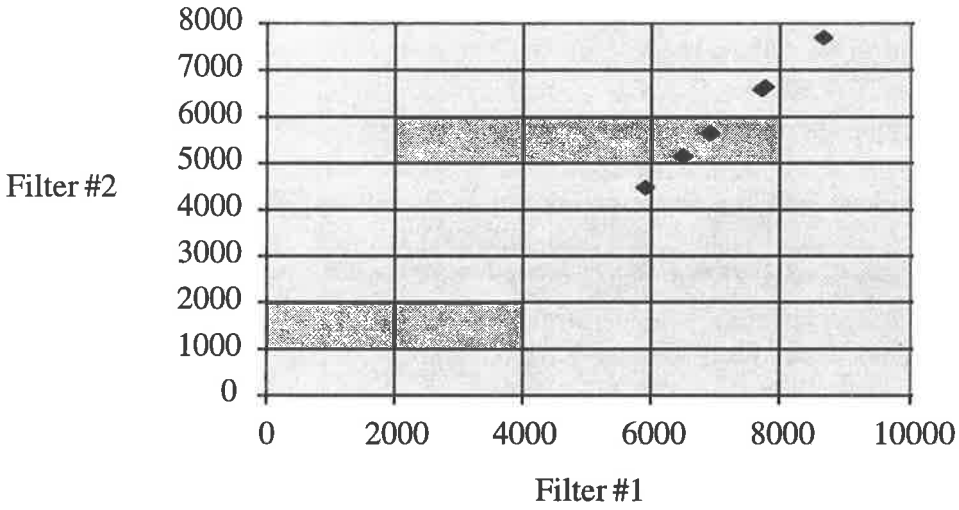


Figure 6.10: Correlation of rRNA gene number with cotton fibre length

a. Correlation of estimated rRNA gene number in six cotton varieties with final fibre length. Experimental values for fibre length were used for the varieties Siokra L-22, Siokra 1-4, CS7S and Strumica 5086, whilst published data was used for Pima S-6 and Paymaster 909, as appears in Table 3.3.

b. Correlation of rRNA gene number estimates between the two filters.

work and that total DNA content does not vary greatly among individuals. The rRNA gene numbers obtained in this experiment are between the previous estimates of Walbot and Dure (1976), Baker *et al.*, (1995) and Wendel *et al.* (1995). The rRNA gene copy number was estimated by Walbot and Dure (1976) in the variety Coker 201 (using hybridisation and reassociation kinetics) at 300-350 copies per haploid genome, and Baker *et al.* (1995) arrived at a similar value for *G. hirsutum* using hybridisation of genomic clones to a heterologous rDNA probe. Our estimate ranges from around 1300 to 2000 copies per haploid genome, whilst the estimate of Wendel *et al.* (1995) is higher, at approximately 3800 repeats per haploid genome. These primary differences are most likely attributable to different estimates used for the size of the haploid genome of cotton.

6.5 Summary and Conclusions

This Chapter describes the results of a preliminary investigation into control of ribosomal RNA synthesis in the developing cotton fibre. A range of cotton varieties reported to show varying fibre lengths were cultivated for examination of their fibre cell length and nucleolar size. Fibre length was measured in four varieties, two of which, Siokra L-22 and Strumica 5086, differed significantly in their final fibre lengths. Nucleolar analysis was carried out on three Australian cotton cultivars in an attempt to verify previous work which is suggestive of a correlation between nucleolar volume and final fibre length. The three *G. hirsutum* varieties under investigation, Strumica 5086, Siokra L-22 and CS7S, are characterised by different nucleolar growth profiles and the time of onset of increase and decrease in nucleolar size differed between varieties with differing fibre length. However, individual nucleoli could not be correlated with particular fibres and a positive correlation (or otherwise) between maximal nucleolar volume and final fibre length could not be inferred from the experiment.

The number of rRNA genes in six cotton cultivars was estimated by comparative slotblots. Significant differences were observed between several varieties, suggestive of

intraspecific variation in rRNA gene number in cotton, but the values obtained did not correlate with final fibre length. The absolute values for rDNA copy number are within the range of previous estimates.

Interestingly, Siokra L-22 and Strumica 5086, which differ in their fibre lengths, also differ in their nucleolar volume profiles. They do not however differ in their numbers of rRNA repeats and it may be differential rRNA expression which causes the differences in nucleolar profiles. Transcription may differ between the varieties and may be correlated with the nucleolar size at various stages of fibre development.

CHAPTER 7

Concluding Discussion

7.1 Analysis of Gene Expression During Cotton Fibre Development

Cotton fibres are differentiated from single ovule epidermal cells and initiate growth at about the time of anthesis, affirming a precise synchrony and homogeneity of growth during their development in cotton bolls. The increasing consumer demand for quality and performance translates into new demands on cotton as a fibre, but prior to the use of genetic manipulation as a tool in cotton fibre improvement, an understanding of the molecular and genetic mechanisms which control commercially important characteristics is required. The work presented in this thesis was primarily aimed at the isolation and characterisation of genes that are predominantly expressed in the fibre cell, under the assumption that some of these genes will be specifically related to the development of the fibre.

7.1.1 Complementary DNA Clones to Fibre-specific mRNA Species

The differential screening of a cDNA library derived from the fibres of an Australian commercial cultivar *Gossypium hirsutum* cv. Siokra 1-2 resulted in the isolation of five fibre-specific messenger RNAs as cDNA clones. Transcript accumulation varied between clones, suggesting that expression of fibre-specific genes is under temporal and spatial regulation coupled with fibre cell differentiation. Screening of the cDNA library with each of the fibre-specific cDNA clones showed that all five fibre-specific transcripts are abundant in elongating fibre cells.

Sequencing of three of the clones revealed significant similarity to known plant proteins. The sequence of pFS1 showed significant sequence similarity to a previously characterised fibre-specific cDNA, E6. The function of the protein encoded by E6 and pFS1 is unknown but it has been suggested to have a unique structural or enzymatic role in cotton fibre primary cell wall deposition.

Clone cFSltp, representing the most plentiful class of fibre-specific cDNA, was identified as encoding a lipid transfer protein (LTP) based on its similarity to LTP mRNAs and genes from a variety of plants. The fibre-specific LTP cDNA sequences fell into two classes. One of

twelve clones sequenced was virtually identical to a fibre-specific LTP isolated from another cotton cultivar, whilst the remaining 11 clones defined a new LTP gene subfamily. LTPs identified from both *Arabidopsis* and broccoli have been located in the cell wall and, in particular, to the waxy cuticle where insoluble cutin may prevent water loss from epidermal cells and provide a mechanical barrier to pathogen ingress. The induction of an analogous activity during cotton fibre development may be related to the extensive cutin deposition which occurs during the elongation phase of growth and is consistent with an abundance of LTP transcripts early in fibre development.

A full-length pFS17 sequence was constructed from the original pFS17 clone sequence and a consensus sequence derived from three 5'-RACE clones. The 1100 nucleotides contained a long open reading frame, potentially encoding a polypeptide of 299 residues, FSPRP, which showed significant sequence similarity to a family of plant proline-rich proteins (PRPs). The precise role of PRPs in plants is unknown but, like LTPs, they are targeted to the cell wall and are thought to have a cross-linking or defence-related function. The primary structure, developmental regulation and tissue specificity of FSPRP suggests that it may have a function in the primary wall deposition phase of cotton fibre development.

Screening of databases with sequences of the remaining two clones, pFS3 and pFS18, was uninformative, implying that they may encode novel protein products. Despite the sequencing of several 5'-RACE clones, it may be that the full-length pFS3 cDNA has yet to be isolated. The persistence of pFS3 transcripts in fibres aged from 6 DPA to at least 20 DPA is suggestive of a possible role for the encoded protein in either elongation or secondary wall deposition or both.

7.1.2 Ribosomal RNA in Cotton Fibres

Studies have shown that ribosomal RNA metabolism may be related to cotton fibre development and a preliminary investigation was carried out of rRNA genes in several Australian cotton cultivars. The nucleolar size in fibres from three *G. hirsutum* varieties,

differing in their final fibre lengths, was measured at early stages of growth. The mean nucleolar diameter reached a maximum at different times after anthesis and declined gradually, confirming previous work. One of several levels at which accumulation of rRNA may be controlled is by quantitative variation in the number of copies of the rRNA gene. The rDNA repeat unit from *G. hirsutum* var. Deltapine 90 was cloned and used to estimate the number of rRNA gene copies in six cotton varieties. Significant differences were observed between some of the varieties, but these did not correlate with final fibre length.

The rate of rRNA gene transcription and the efficiency of precursor processing have not been addressed in this project. Future work might involve a direct investigation of the number of active transcription complexes by nucleolar dispersal and silver staining (Haaf *et al.*, 1991). In addition, biosynthesis and efficiency of processing of pre-rRNA could be determined by determination of the specific activity of the ATP pools in the initial and processed products of transcription (Ni Ghogain *et al.*, 1982) and the half-life of the ribosomes determined by pulse-chase or continuous labelling experiments. Several *trans*-acting factors involved in RNA polymerase I transcription initiation have been identified in vertebrates (Reeder, 1990) and an investigation of plant homologues may also be fruitful.

7.2 Future Prospects for Genetic Engineering of Cotton

Knowledge gained from this work contributes to our understanding of cotton fibre development and may have significant agricultural and commercial ramifications for the Australian cotton industry. Future work should focus on the obtaining of promoters which correspond to the fibre-specific cDNAs described in this study. Two members of the LTP gene family have thus far been identified in the genome of Siokra 1-2, but it appears unlikely that either encodes the fibre-specific transcripts of interest.

The first step in the evaluation of fibre-specific promoters is isolation of the corresponding genes followed by verification of tissue-specific expression patterns by transient

expression in cotton plants. Previous experiments with cotton gene promoters have utilised promoter sequences of 2.5 kb (John and Crow, 1992) and 2.3 kb (John, 1995), but as little as 101 bp was sufficient to direct anther-specific expression in transgenic tobacco (Koltunow *et al.*, 1990). Several hundred bp of the immediate 5' flanking regions of plant genes are generally sufficient to confer the appropriate pattern of expression on heterologous reporter genes (Willmitzer, 1988). Transformation techniques based on particle bombardment allow epidermally transformed plants (McCabe and Martinell, 1993), which could provide useful data for the evaluation of fibre-altering genes such as the ones isolated in this study.

Cotton crops are extremely vulnerable to bacterial and insect pests, some of which specifically attack the cotton boll. Development of transgenic plants which express resistance genes in susceptible regions will result in a reduction in the massive chemical input necessary for efficient cotton production. Foreign genes, potentially from any source, may be transferred into cotton plants and accurately tissue-targeted to the fibres without detrimental effects on overall plant vigour. In addition, the genes identified in this study are expressed at distinct times in fibre development and their temporal expression patterns could be exploited in transgenics. Promoters which drive pFS6, pFS17 and pFS1 gene expression could be used to express heterologous genes early in fibre development, during cell elongation. pFS18 expression, on the other hand, peaks at 12-14 DPA whilst pFS3 transcripts were detected from 6 to 20 DPA.

Future research efforts will also aim to determine whether there is a common *cis* regulatory element in fibre-specific promoters which is sufficient to direct tissue-specific transcription. Manipulation of such control regions may result in the specific alteration of the sequential progression of several genes or the timing and/or level of expression of a particular gene. A gene's structure might also be modified to give novel protein products. The ability to decrease the expression of a gene in a transgenic plant has potential utility in the study of plant gene expression and function as well as in crop improvement and offers a method for

constructing a mutant.

Transferral of intact genes into heterologous plant hosts has been the subject of a number of experiments (Willmitzer, 1988). Future work might also involve transferral of the intact LTP gene into a heterologous host such as *Arabidopsis* to determine whether the tissue-specific expression is retained, perhaps in the trichomes. In addition, it would be of interest to ascertain if the expression patterns of fibre-specific genes such as PRPs and LTPs differ between cultivars which vary in their fibre characteristics or between varieties with wild type and mutant fibres. A drawback to the use of differential screening techniques is that they only allow isolation of abundant mRNAs. It may be that the crucial proteins in fibre development are rarer regulatory proteins and may be identified by techniques such as differential display. Characterisation of "cold" clones from a fibre cDNA library may also prove fruitful in this respect.

7.3 Final Statements

Many of the aims outlined prior to the commencement of this research have been achieved and results from this project provide a foundation for future work on fibre-specific promoters. Further characterisation of the clones described here and isolation of additional genes may help in gaining further insight into the developmental processes which occur in elongating cotton fibres, and will eventually allow breeders to respond rapidly to changes in spinning technology and to increase the rate of fibre improvement more generally.

APPENDICES

APPENDIX 1

Cotton fibre measurements: *G. hirsutum*, cv. Siokra L-22

Cotton plant	Fruiting branch (approx.)	Boll age (DPA)	Fibre length of 6 seeds (mm)						Ave. length for boll (mm)	St. dev. for boll (mm)
4	19	5	2	3.5	3	2	2	1.5	2.33	0.75
3	19	5	3	2.5	2	3	2	2	2.42	0.49
4	20	5	2	2.5	2	2.5	3	2	2.33	0.41
5	19	5	3	3	2	2.5	3	2	2.58	0.49
8	2	5	1.5	2	2	1.5	1.5	2	1.75	0.27
8	4	5	2	2	3	2	3	2	2.33	0.52
4	14	10	10	10.5	10	10.5	8	9	9.67	0.98
5	16	10	11	12	10	10	10	9	10.33	1.03
7	3	10	11	9.5	10	10	9	9	9.75	0.76
5	3	10	7	7	6	8	4	8	6.67	1.51
3	13	15	19	21	22	19	22	21	20.67	1.37
5	13	15	24	20	21	22	21	21	21.50	1.38
4	10	15	25	22	22	25	24	24	23.67	1.37
3	2	15	14	18	16	15	16	15	15.67	1.37
7	2	15	22.5	24	24	23	23	22.5	23.17	0.68
5	2	15	18	19	16	17	15	18	17.17	1.47
2	14	20	28	25.5	25	27	26	26	26.25	1.08
2	17	20	31	32	30	29	31	29	30.33	1.21
2	4	20	25.5	26	27	29	26	29	27.08	1.56
1	4	20	26	27	26	27	27	29	27.00	1.10
1	10	20	26	26	25	24	23	24	24.67	1.21
4	9	25	26	31	26	28	32	27	28.33	2.58
5	8	25	30	28	29	30	28	29	29.00	0.89
1	6	25	30	28	31	32	31	32	30.67	1.51
4	2	25	31	30	28	30	28	29	29.33	1.21
3	13	30	30	31	30	31.5	31	32	30.92	0.80
2	2	30	30	29	29	32	29	31	30.00	1.26
2	3	30	26	31	30	29	28	30	29.00	1.79
1	2	30	29	27	28	30	27	28	28.17	1.17
7	1	30	27	30	28	28	29	30	28.67	1.21
4	8	35	33	30	31	31	31	32	31.33	1.03
5	9	35	31	30	30	28	29	30	29.67	1.03
6	2	35	31	32	30	28	30	32	30.50	1.52
2	2	35	26	29	29	29.5	28	30	28.58	1.43
4	1	35	30	32	31	29	30	30	30.33	1.03
1	2	40	27	25	26	28	26	28	26.67	1.21
5	5	40	28	26	28	29	28	30	28.17	1.33
1	2	40	31	32	31	32	30	30	31.00	0.89
1	5	40	31	31	32	30	30	31	30.83	0.75
2	10	40	27	30	29	28	27	31	28.67	1.63
5	4	45	31	27	29.5	30	27	29	28.92	1.63
4	4	45	31	29	27	28	28	26	28.17	1.72
6	1	45	29	32	30	31	28	31	30.17	1.47
3	2	45	32	31	30	30	31	32	31.00	0.89
3	8	45	28	30	30	31	29	29	29.50	1.05
2	2	50	32	31.5	29.5	31	31	30	30.83	0.93
3	2	50	28	29	31	30	27	30	29.17	1.47
1	1	50	32	29	31	31	28	31	30.33	1.51
3	1	50	31	30	31	29	29	28	29.67	1.21
2	6	50	31	28	30	29	29	31	29.67	1.21

APPENDIX 2

Cotton fibre measurements: *G. hirsutum*, cv. Siokra 1-4

Cotton plant	Fruiting branch (approx.)	Boll age (DPA)	Fibre length of 6 seeds (mm)						Ave. length for boll (mm)	St. dev. for boll (mm)
1	11	5	3	1.5	2.5	1.5	2	2	2.08	0.58
1	9	5	2	2.5	3	2	2.5	2.5	2.42	0.38
2	9	5	2	2	3	2	2.5	2.5	2.33	0.41
1	15	10	9	12	10.5	9	10	10	10.08	1.11
1	19	10	9	8	9	7	9	7.5	8.25	0.88
5	22	10	6	10	10	9	9	9	8.83	1.47
3	13	10	9	10.5	9	10	8	11	9.58	1.11
1	17	15	19	18	18	20	20	16	18.50	1.52
5	15	15	19	14	21	21	18	20	18.83	2.64
3	16	15	25	24	22	25	26	15	22.83	4.07
5	18	20	30	28	30	29	29	29	29.17	0.75
5	19	20	25	29	25.5	30	28	28	27.58	1.96
2	14	20	27	30	30	32	30	29	29.67	1.63
5	13	20	30	30.5	29	29	28	30	29.42	0.92
2	10	25	26	24	27	26.5	30	27	26.75	1.94
4	9	25	26	27	27	27	26	26	26.50	0.55
2	4	25	28	28	26	26	28	29	27.50	1.22
2	5	25	25	28	28	25	27	27	26.67	1.37
4	8	30	28	25	28.5	28	29.5	29	28.00	1.58
1	15	30	28	29	27	28	29	29	28.33	0.82
3	5	30	28	28	27	28.5	27	29	27.92	0.80
5	10	35	27	29	28.5	28	30	29	28.58	1.02
3	8	35	28	30	31	31	30	29	29.83	1.17
2	1	35	27	27	28	26	27	26	26.83	0.75
2	3	35	27	26	24	27	28	25	26.17	1.47
5	3	40	30	31	28	30	31	31	30.17	1.17
1	2	40	30	28	26	27	29	27	27.83	1.47
1	3	40	27	29	28	28	28	30	28.33	1.03
3	4	40	31	28	30	28	28	29	29.00	1.26
5	3	45	28	31	31	29	29	29	29.50	1.22
2	11	45	27	29	31	27	28	28	28.33	1.51
1	1	45	27	28	29	30	29	29	28.67	1.03
3	3	45	27	29	28	30	28	29	28.50	1.05
1	7	50	28.5	29	30	30.5	31.5	30	29.92	1.07
1	7	50	27	26	26	25	24.5	25	25.58	0.92
3	2	50	27	27	29	28	27	26	27.33	1.03
3	1	50	27	28	28	30	28	28	28.17	0.98

APPENDIX 3

Cotton fibre measurements: *G. hirsutum* cv., CS7S

Cotton plant	Fruiting branch (approx.)	Boll age (DPA)	Fibre length of 6 seeds (mm)						Ave. length for boll (mm)	St. dev. for boll (mm)
5	15	5	3	2	2	1.5	2.5	2	2.17	0.52
2	13	5	2.5	2	2	2	2	2.5	2.17	0.26
2	10	5	2	1.5	2	2.5	2	2	2.00	0.32
5	3	5	2	2.5	2	2	2.5	2	2.17	0.26
2	18	10	11	10	10	11.5	10	12	10.75	0.88
2	15	10	12.5	13	12	10	11	12	11.75	1.08
5	16	10	9	10	11	8	9	8	9.17	1.17
2	11	10	9	10	11	11	11	9	10.17	0.98
2	3	10	8	10.5	9	8.5	10	8	9.00	1.05
5	16	15	18	25	22	22	22	19	21.33	2.50
2	12	15	24	25	23	21	23	22	23.00	1.41
2	13	15	23	21	22	24	22	23	22.50	1.05
4	3	15	21	21	18	18	17.5	18	18.92	1.63
8	1	15	16	18	18	17	17	19	17.50	1.05
2	10	20	26	26	27	25	25	26	25.83	0.75
2	13	20	26	24	25	25	25	26	25.17	0.75
2	11	20	25	23.5	25	24	22	23	23.75	1.17
3	5	20	25	23	27	26	23	24	24.67	1.63
3	2	20	26	24	23	26	25	25	24.83	1.17
4	4	20	25	25	27	23	26	24	25.00	1.41
1	4	25	27	24	26	22	21	25	24.17	2.32
4	4	25	26	23	23.5	24	26	22.5	24.17	1.51
3	4	25	23	26	25	25	28	24	25.17	1.72
6	4	25	28	27	26	26	26	24	26.17	1.33
3	1	25	25	24	28	25.5	25	26	25.58	1.36
1	4	30	29	26.5	24	25	27	25	26.08	1.80
4	8	30	28	27	28	26	27	25	26.83	1.17
1	6	30	26	25	25	24	23	26	24.83	1.17
4	2	30	27	25	24	25	24	26	25.17	1.17
7	1	30	28	28	26	27	28	27	27.33	0.82
5	3	35	29	29	29	27	27	27	28.00	1.10
1	1	35	28	30	30	28.5	30	28	29.08	1.02
1	5	35	26	25	25	28	26	25	25.83	1.17
3	3	35	25	26	27	26	27	25	26.00	0.89
2	1	35	26	26	25	27	26	25	25.83	0.75
4	5	35	29	27	25	27	27	25	26.67	1.51
2	2	40	24	28.5	29	26.5	24	25.5	26.25	2.16
3	3	40	29	27	25	29.5	27	28	27.58	1.63
1	4	40	26	25	26	28	26	25	26.00	1.10
2	4	40	26	27	24	25	25	26	25.50	1.05
2	5	40	26	25	25	27	25	25	25.50	0.84
5	5	40	24	25	24	25	26	25	24.83	0.75
5	1	45	24	26	24	25.5	28	25	25.42	1.50
5	2	45	28	23	27.5	28	26	26	26.42	1.91
1	3	45	27	23	23	26	26	25	25.00	1.67
5	2	45	25	26	26	25	25	26	25.50	0.55
7	4	45	27	27	26	27	26	26	26.50	0.55
2	2	50	25	28	29	23	28	26	26.50	2.26
3	1	50	27	25	26	29	29	24	26.67	2.07
1	2	50	19	28	27	27	26	26	25.50	3.27
1	2	50	26	25	25.5	26	26	24	25.42	0.80
1	1	50	25	25	27	26	26	25	25.67	0.82

APPENDIX 4

Cotton fibre measurements: *G. hirsutum*, cv. Strumica 5086

Cotton plant	Fruiting branch (approx.)	Boll age (DPA)	Fibre length of 6 seeds (mm)							Ave. length for boll (mm)	St. dev. for boll (mm)
4	?	5	3	3	2	2	2.5	2		2.42	0.49
2	?	5	2	3	2	2	2	2		2.17	0.41
5	4	5	2	1.5	2	2	2	1.5		1.83	0.26
2	?	5	2.5	2	2.5	2	2	2.5		2.25	0.27
2	18	10	11	10	10	11.5	10	12		10.75	0.88
2	15	10	12.5	13	12	10	11	12		11.75	1.08
5	16	10	9	10	11	8	9	8		9.17	1.17
2	11	10	9	10	11	11	11	9		10.17	0.98
2	3	10	8	10.5	9	8.5	10	8		9.00	1.05
1	16	15	19	21	21	20	21	22		20.67	1.03
2	6	15	17	16	15	16	17	17		16.33	0.82
3	2	15	18	19	18	19	19	20		18.83	0.75
7	6	15	19	19	17.5	18	20	18		18.58	0.92
5	1	15	15	17	19	18	17	17		17.17	1.33
2	7	20	24	21	24	23	25	22		23.17	1.47
2	13	20	27	25	22	26	23	27		25.00	2.10
1	14	20	23	22	22	24	23	22		22.67	0.82
6	4	20	22	21.5	24	22	22	21		22.08	1.02
5	6	20	21	20.5	22	21	22.5	22		21.50	0.77
4	10	25	21	24	22	21	24	22		22.33	1.37
4	4	25	21	22	22	21	23	23		22.00	0.89
4	6	25	21	23	22	22	21	24		22.17	1.17
5	3	25	24	21	22	20	21	21		21.50	1.38
3	8	25	21	22	22	23	22	20		21.67	1.03
3	2	30	23	24	20	25	22	22		22.67	1.75
2	5	30	25	25	24	24	23	24.5		24.25	0.76
2	13	30	31	30	26	26	27	26		27.67	2.25
3	4	30	23	20	19	18	21	20		20.17	1.72
8	6	30	22	21	23	22.5	21	21		21.75	0.88
4	8	35	23	22	23	24	23.5	23		23.08	0.66
1	8	35	25	24	24.5	25	25	22		24.25	1.17
1	10	35	19	24	21	22	23	23		22.00	1.79
6	6	35	22	24	19	22	23	22		22.00	1.67
6	8	35	25	25	22	21	21	22		22.67	1.86
4		40	22	22	23	21.5	23	19		21.75	1.47
2	2	40	27	24	26	27	26	25		25.83	1.17
1	4	40	23	22	21.5	22	22	21.5		22.00	0.55
1	4	40	21	18.5	21	20.5	21	22		20.67	1.17
4	6	40	20	23	22	26	22	22		22.50	1.97
1	3	45	24	25	24	26	22	23		24.00	1.41
3	1	45	21	20	22	22.5	20	23		21.42	1.28
2	1	45	25	26	25	24	27	24		25.17	1.17
2	3	45	22	21	20	21.5	22	22		21.42	0.80
3	6	45	22	23	24	23	21	21		22.33	1.21
1	1	50	26	27	27	28	26	26		26.67	0.82
1	4	50	27	24	21	25	23	25		24.17	2.04
1	1	50	25	24	21	22	22	23		22.83	1.47
1	3	50	21	21	22	21	23	20		21.33	1.03
7	5	50	23	21	22	21	22	21		21.67	0.82

APPENDIX 5

Nucleolar diameter measurements: *G. hirsutum*, cv. Siokra L-22

Fibre age (DPA)	2				4				6				8				10				12				14	16
Seed	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**1**
Diameter (µm)	4.9	5.78	5.5	5.67	5.08	6.02	6.73	5.08	7.8	7.33	9.14	8.14	6.01	6.34	6.03	5.92	6.07	6.93	7.6	5.41	6.36	5.84	5.78	5.9	4.6	6.26
	6.3	5.22	5.23	4.83	6.18	5.63	6.01	6.18	7.45	6.58	9.13	8.08	6.03	5.92	6.68	6.46	7.12	7.16	8.06	5.14	5.96	5.69	6.12	5.64	5.11	5.03
	5.17	4.7	4.87	6.18	4.64	6.1	6.84	4.64	7.83	7.98	6.39	7.63	6.19	6.21	6	5.81	7.51	7.02	6.6	4.93	6.53	6.4	6.08	5.57	5.52	6.46
	4.46	5.11	5.96	5.07	4.8	5.97	6.63	4.8	8.07	8.03	8.51	8.13	5.76	5.94	6.73	6.38	7.31	8.08	7.81	5.36	6.83	6.37	6.45	6	5.03	5.17
	5.29	5.44	5.74	5.78	5.37	5.62	5.64	5.37	7.86	6.95	7.82	6.81	5.85	5.58	5.84	7.39	7.08	6.36	6.69	5.11	6.08	6.61	5.97	6.02	5.12	6.39
	5.14	5.99	4.93	5.15	4.65	6.16	6.24	4.65	7.06	6.62	8.9	6.16	5.7	6.27	7	6.72	5.91	6.87	6.98	4.74	6.84	5.96	6.21	6.92	5.43	5.76
	4.79	6.25	4.7	5.17	4.98	4.73	6.7	4.98	7.74	8.17	6.87	8.11	6.22	5.24	6.02	6.41	7.11	6.14	6.68	5.83	6.55	6.59	6.47	5.79	4.52	6.17
	4.79	5.58	6.23	5.58	4.64	5.55	6.69	4.64	8.07	8.05	7.08	7.25	6.07	5.37	5.96	7.01	6.66	7.92	6.8	4.56	6.84	6.16	5.86	5.25	5.8	5.38
	4.8	5.24	5.24	5.05	5.79	6.1	6.97	5.79	7.09	6.95	6.12	6.72	6.45	4.92	6.21	6.95	7.56	7.43	6.38	5.79	5.81	6.37	6.39	7.09	5.19	4.48
	6.14	5.72	6.2	5.22	4.3	5.62	6.31	4.3	7.39	7.46	7.2	8.41	5.77	4.45	6.49	6.38	6.87	6.29	7.15	5.62	5.66	6.16	6.5	7.2	4.67	5.27
	5.17	5.4	5.93	4.31	4.7	6.13	6.56	4.7	7.35	7.4	8.47	7.23	5.76	4.85	6.5	6.41	7.26	7.06	7.2	5.71	5.65	5.52	6.48	5.78	5.82	4.83
	5.8	5.28	6.74	5.17	4.78	6.71	6.73	4.78	7.39	8.12	8.63	8.01	6.12	4.73	6.01	5.2	7	6.56	6.64	6.09	5.54	6.74	6.3	6.58	5.16	4.9
	4.99	5.41	6.75	5.17	4.64	6.13	7.46	4.64	7.61	8.05	6.48	7.73	5.98	5.3	6.57	5.84	7.05	7.23	6.64	5.7	6.09	6.28	6.16	6.7	5.55	4.27
	4.84	5.27	6.67	5.14	5.66	5.5	6.92	5.66	7.45	8.01	7.63	6.89	5.58	5.03	6.7	6.12	6.86	7.25	7.07	5.87	5.44	6.16	5.91	6.59	5.59	5.04
	5.66	5.32	5.15	5.51	4.82	5.75	6.56	4.82	7.08	7.35	7.8	7.56	6.89	5.78	6.82	5.42	6.74	7.89	7.08	6.08	6.64	5.6	5.98	6.2	6.35	4.44
	5.1	6.1	6.18	6.17	4.26	5.47	6.89	4.26	7.93	6.92	7.74	8.24	6.95	5.35	6.14	5.44	7.48	7.69	7.67	5.98	5.95	5.43	6.76	6.74	5.47	5.45
	4.34	5.88	5.02	5.15	4.38	6.18	6.77	4.38	6.89	8.13	8.09	6.67	6	5.01	5.77	5.82	6.12	6.99	7	5.18	6.25	6.32	6.11	8.72	6.09	4.89
	4.71	5.89	6.86	5.73	4.07	6.27	7.05	4.07	7.01	6.67	8.02	6.74	6.6	5.3	6.15	5.84	6.74	7.68	7.16	4.85	6.22	6.01	6.36	5.51	4.44	4.39
	5.2	5.94	5.65	4.01	4.99	6.18	6.89	4.99	7.63	7.67	7.03	7.41	6.37	5.76	5.58	7.28	6.35	6.46	7.07	6.03	6.27	7.53	6.34	5.31	5.15	5.55
	4.57	6.15	4.71	5.52	4.7	6.21	6.49	4.7	7.7	7.26	7.32	7.72	6.11	4.42	6.4	6.03	7.19	6.98	6.72	5.8	6.79	7.15	6.73	4.35	4.79	5.06
	4.54	5.65	6.25	5.94	4.6	6.38	6.26	4.6	6.73	6.86	7.51	7.81	6.47	5.05	6.68	6.16	7.03	7.55	6.47	5.96	6.61	6.5	6.33	6.36	4.97	5.58
	4.78	5.22	5.9	5.9	4.84	6.1	7.06	4.84	7.23	7.24	6.62	8.18	5.86	4.81	8.24	6.54	6.98	7.9	7.17	5.33	6.34	7.31	6.54	7.26	5.74	5.55
	5.25	5.08	6.77	4.85	4.65	6.82	6.68	4.65	6.52	7.25	6.82	5.96	5.97	5.1	6.84	6.69	6.3	4.7	6.77	6.02	5.69	6.37	6.24	6.02	5.44	4.8
	5.63	5.56	6.67	5.72	5.97	6.18	6.85	5.97	6.47	6.76	7.33	8.09	5.87	5.49	6.94	6.33	6.73	6.84	6.72	6.07	6.22	6.87	6.37	7.84	5.03	4.94
	6.04	5.8	6.14	4.96	6.83	5.57	6.42	6.83	6.09	6.64	7.69	8.11	6.32	5.48	7.15	6.21	5.48	7.57	7.07	5.85	7	6.81	6.58	6.39	5.75	5.19

APPENDIX 6

Nucleolar diameter measurements: *G. hirsutum*, cv. CS7S

Fibre age (DPA)	2				4				6				8				10				12				14
Seed	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**
Diameter (μm)	6.03	5.59	6.13	6.03	5.65	6.42	5.53	5.82	6.93	5.33	5.3	5.2	5.35	6.26	5.71	5.55	7.47	5.3	6.27	5.92	6.19	7.12	7.65	5.6	5.58
	7.28	5.65	6.35	6.95	5.6	6.33	6.34	6.06	7.35	5.55	5.78	7.05	5.44	5.73	5.17	5.6	5.49	5.18	4.93	5.61	5.95	6.99	8.14	5.68	5.25
	6.71	6.41	5.94	7.04	5.62	5.55	5.89	6.04	7.1	5.14	5.16	7.07	5.54	6.06	5.54	7	6.07	6.29	5.74	5.9	5.85	6.52	7.27	5.57	6.53
	6.41	5.86	5.97	6.06	5.02	6.31	6.26	5.74	7.18	5.63	5.53	5.94	5.36	5.89	5.43	6.51	6.74	5.31	5.46	5.15	5.93	6.83	7.37	6.12	5.87
	7.25	5.87	5.78	6.03	5.88	6.42	5.6	6.86	8.39	5.21	5.03	7.01	5.63	6.27	4.71	6.56	6.12	5.35	5.25	6.69	6.18	6.57	6.02	5.97	5.77
	7.21	5.86	6.49	5.89	5.48	6.74	5.61	5.77	6.4	5.02	5.28	5.76	5.45	5.76	5.65	6.68	6.12	5.54	5.37	5.67	6.39	6.7	7.96	6.04	6.35
	6.31	6.23	6.35	6.31	5.24	6.26	6.28	6.01	7.15	5.32	5.98	6.16	5.14	5.6	5.45	7.4	6.81	5.7	5.75	5.58	6.57	7.22	5.35	6	5.77
	6	5.46	6.7	6.01	6.67	6.25	6.43	6.51	6.49	6.18	5.5	6.41	5.82	5.51	5.76	7.61	6.04	5.68	5.27	5.45	5.54	7.54	5.6	6	5.35
	7.34	6.44	6.61	5.84	5.18	5.96	6.89	5.69	6.79	5.46	5.41	6.53	5.44	6.05	5.45	6.75	5.81	5.08	5.62	6.11	6.88	6.28	6.19	6.44	5.96
	7.2	5.54	6.31	5.94	5.44	6.48	6.12	5.83	6.22	5.41	5.77	6.42	5.45	5.89	5.3	7.11	5.67	5.95	5.72	6.06	7.06	7.41	6.76	5.53	6.03
	7.58	6	6.51	6.04	5.42	5.41	5.81	6.05	6.62	5.24	5.55	6.32	5.05	5.99	5.08	5.97	6.57	6.13	5.1	5.3	6.21	7.63	7.03	6.05	5.64
	7.1	6.35	6.53	6.08	5.53	6.06	5.92	5.76	6.92	5.79	6.11	6.42	5.35	5.88	5.91	6.4	6.09	5.78	5.31	5.9	6.66	7.55	6.11	6.04	5.89
	7.84	6.69	6.04	7.21	5.45	6.39	5.91	5.82	7.11	5.24	5.94	6.81	5.63	5.57	5.83	7.55	6.1	5.69	5.46	5.36	6.18	7.18	6.29	5.21	6.17
	6.62	6.01	6.36	6.87	5.56	6.92	5.56	5	7	5.84	5.54	6.61	5.39	5.62	5.44	6.15	6.68	5.59	5.57	5.06	6.72	7.29	5.41	5.72	6.46
	6.7	5.28	6.24	6.54	5.62	6.27	5.74	5.76	7.12	5.67	5.03	6.94	5.66	6.06	6.06	6.11	5.67	5.77	5.27	5.3	6	7.89	6.37	5.17	5.7
	5.89	6.15	6.65	6	6.46	6.41	6.03	5.96	7.25	5.74	5.73	7.09	5.66	5.66	5.53	6.49	6.23	5.19	5.48	5.49	6.2	6.7	6.57	6.39	5.98
	6.36	5.81	6.78	6.02	5.45	7.02	5.56	5.83	7.2	5.86	5.6	6.22	5.28	5.54	5.83	6.83	5.31	5.37	5.63	5.78	6.88	7.63	5.87	6.34	6.15
	5.96	5.55	6.6	6.01	5.61	6.67	6.33	6.12	7.46	6.22	5.55	6.42	5.71	5.86	5.47	7.34	6.35	5.92	5.41	5.57	6.66	7.57	6.3	5.89	5.68
	6.33	5.38	6.47	6.52	6.02	5.81	6.77	5.88	6.8	6.38	5.17	7.58	5.7	5.47	5.85	6.55	7.39	5.66	5.64	5.57	6.28	6.59	6.14	5.99	6.19
	5.81	5.7	6.79	6.23	5.56	5.78	6.99	6.07	7.46	5.22	5.33	6.27	6.3	6.05	5.22	5.41	5.84	5.13	5.49	5.52	6.99	6.31	6.47	5.78	5.8
	6.16	7.21	6.66	6.78	6.21	6.91	6.66	5.96	6.91	5.62	4.82	6.73	6.32	5.51	6.33	6.17	6.33	5.89	7.28	5.79	6.38	6.2	7.01	5.54	5.69
	6.01	6.34	5.98	6.49	5.56	6.44	6.21	6.58	6.33	5.94	5.46	6.43	5.16	5.33	4.92	6.66	5.64	5.8	5.36	6.3	6.03	7.31	5.72	5.76	6.08
	6.26	6.31	5.63	6.38	5.89	6.44	6.06	5.44	6.59	5.77	5.06	6.34	5.03	5.4	5.72	5.99	5.7	5.78	5.56	5.42	6.83	6.51	5.76	5.73	5.82
	6.42	6.48	7.08	7.25	6.54	6.31	5.7	6.56	6.61	6.15	5.62	6.12	5.55	6.25	5.08	7.09	5.24	5.82	6.41	5.59	6.39	6.42	6.21	5.47	6.22
	5.81	6.6	6.29	6.38	5.5	6.24	6.12	6.03	6.21	4.85	5.26	5.91	5.79	5.58	6.31	6.17	5.21	5.41	6.75	6.06	6.04	6.26	5.33	5.74	6.92

APPENDIX 7

Nucleolar diameter measurements: *G. hirsutum*, cv. Strumica 5086

Fibre age (DPA)	2				4				6				8				10				12				14
Seed	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**
Diameter (µm)	6.93	7.07	6.52	6.71	9.28	9.03	6.76	8.23	7.46	6.9	6.29	8.1	4.36	6.56	6.56	5.3	5.88	5.87	5.13	6.36	6.52	6.08	4.79	6.23	5.12
	7.28	7.12	6.91	6.81	8.09	8.36	7.78	7.33	7.51	7.3	6.74	8.46	6.29	6.45	6.52	5.32	6.08	5.82	5.3	5.55	5.76	5.61	4.54	5.73	5.75
	7.44	7.83	6.81	6.22	7.78	8.22	7.86	7.75	7.22	6.88	6.98	7.89	6.09	6.73	6.56	6.32	6.25	6.74	5.73	6.04	5.41	5.27	4.95	5.1	5.58
	6.39	6.87	6.62	6.41	9.04	9.17	7.49	7.15	7.95	6.7	6.55	6.71	7.04	6.43	6.39	6.57	6.15	5.99	5.36	5.99	5.63	5.65	4.52	5.71	5.54
	7.59	6.99	6.19	7.42	8.44	10.54	8.3	8.42	7.46	6.91	6.42	7.52	7.16	6	6.29	6.97	6.73	6.07	4.53	6.1	6.39	5.19	5.11	6.96	5.95
	7.19	7.08	6.38	7.37	8.98	9.04	7.48	8.69	7.9	7.13	7.4	7.08	7.55	6.4	6.13	5.64	6.5	6.45	5.77	5.89	5.89	5.33	5.86	5	6.17
	6.69	8.17	6.23	6.46	8.88	8.78	8.1	9.04	7.41	6.94	6.68	6.02	5.99	6.46	5.28	6.36	7.4	6.29	5.19	5.62	5.91	6.04	5.72	6.07	5.22
	7.16	7.88	6.98	6.94	8.62	7.89	7.76	7.79	6.59	7.34	6.83	6.55	6.07	6.24	5.67	6.26	6.57	7.36	5.42	5.71	6.64	5.2	6.45	6.58	5.23
	7.11	6.59	6.92	6.03	8.66	9.19	8.37	7.68	7.13	6.27	6.74	6.95	6.69	6.15	6.43	6.02	6.95	6.44	5.11	5.73	5.38	5.3	5.43	5.43	5.53
	7.16	7.28	7.39	6.02	8.2	9.39	7.84	6.64	8.45	6.57	6.22	8.68	6.36	6.69	6.35	6.31	6.36	5.05	5.49	5.69	4.92	5.6	6.08	6.41	5.52
	7.14	7.22	6.35	6.88	8.29	7.55	7.58	6.99	7.75	6.72	6.8	6.91	6.54	7.05	5.6	5.58	6.35	7.04	5.02	6.83	5.35	5.68	5.75	5.87	5.81
	7.27	6.28	6.18	7.16	6.72	7.49	7.38	7.75	7.87	7.13	6.41	7.74	6.14	6.74	5.51	5.87	6.42	5.65	5.24	6.25	6.09	5.63	5.77	6.38	6.15
	6.83	7.03	6.68	6.79	8.55	7.94	8.21	7.23	7.29	6.48	6.89	7.79	6.52	6.34	5.42	6.11	6.62	6.71	6	6.45	4.79	5.72	7	5.58	6.11
	7.73	8.07	7.12	7.5	8.27	7.6	8.6	6.96	7.13	6.71	6.89	7.68	6.3	7.01	6.15	6.08	6.53	6.71	5.36	6	5.5	5.44	5.4	6.07	5.95
	6.82	7.58	6.36	6.75	8.35	7.69	7.03	7.15	7.37	7.02	7.8	7.86	5.76	6.24	6.03	6.42	7.05	6.57	5.63	5.76	5.07	5.21	5.59	6.48	6.13
	7.24	7.96	6.7	6.21	7.9	6.95	7.18	7.16	7.59	6.95	7.44	7.9	4.82	6.72	5.97	6.4	6.94	6.04	5.64	6.43	6.81	6.97	6.04	5.91	5.74
	7.94	8.2	7.48	7.85	8.62	7.69	7.62	7.88	8.72	7.76	7.64	7.88	6.6	6.33	6.13	7	6	6.34	5.28	6.66	7.15	5.82	5.2	5.67	6.55
	7.47	8.24	7.13	6.41	8.1	8.73	7.59	7.6	8.09	7.55	7.07	8.18	5.8	6.57	6.35	7.1	6.15	6.21	5.66	5.55	5.84	6.56	5.64	5.75	6
	7.39	7.74	6.99	7.12	8.85	7.65	7.29	7.27	7.26	7.23	7.29	8.02	6.26	6.39	5.93	6.27	6.28	6.72	5.56	6.43	5.61	5.75	5.76	5.29	6.28
	6.54	8.74	6.78	7.12	7.65	7.79	7.64	7.22	7.17	8.61	7.24	7.41	5.6	6.16	6.01	6.49	5.78	6.02	5.68	5.48	5.7	5.74	6	5.86	5.41
	7.04	7.3	7.26	7.25	7.27	8.42	6.95	7.12	6.66	7.34	7.44	7.7	5.96	6.66	5.62	7.4	5.84	6.08	5.48	5.82	5.58	6.59	5.65	6.1	5.90
	6.97	7.06	6.22	6.82	8.27	8.01	8.84	7.6	6.55	7.86	8.02	8.35	6.66	6.17	5.56	5.57	7.09	6.37	5.12	6	4.95	6.34	5.77	6.19	5.47
	6.68	7.1	7.33	7.22	9	9.31	8.19	6.22	7.39	6.71	9.06	7.95	6.85	6.96	5.39	6.18	5.8	6.25	5.13	6.89	5.93	5.61	6.26	5.88	6.01
	7.19	7.69	6.35	7.59	7.96	8.06	7.71	7.16	7.34	6.64	7.24	8.16	6.17	6.16	6.01	6.96	5.22	6.66	5.25	6.53	5.28	5.72	5.25	5.27	5.98
	7.11	8.01	6.72	6.91	8.01	8.6	7.54	7.29	7.43	7.16	7.01	6.13	5.75	6.19	5.54	7.06	5.77	6.41	5.51	6.26	5.88	6.58	6.5	7.34	6.47

APPENDIX 8

a. Diphenylamine DNA assays of calf thymus DNA standards

DNA amount (µg)	Data (OD_{600}s)			Average OD_{600}	St. dev.	Line of best fit (trend)
0.0	0.000	0.000	0.000	0.000	0.000	0.017
2.4	0.098	0.132	0.089	0.106	0.023	0.116
4.8	0.234	0.212	0.217	0.221	0.012	0.215
7.2	0.324	0.312	0.324	0.320	0.007	0.314
9.6	0.435	0.430	0.433	0.433	0.003	0.413
12.0	0.566	0.506	0.531	0.534	0.030	0.512
14.4	0.598	0.610	0.620	0.609	0.011	0.611
16.8	0.698	0.701	0.710	0.703	0.006	0.710
19.2	0.766	0.802	0.797	0.788	0.020	0.808

Regression of the standard curve: $y = 0.04121x + 0.01716$

b. Diphenylamine DNA assays of genomic and plasmid DNA samples

Cotton variety	Data (OD_{600}s)			Average OD_{600}	St. dev.	DNA amount (µg/10 µl)	DNA amount (µg/µl) in stock solution	No. µl needed for 1µg
pGhR1	0.182	0.233	0.219	0.211	0.026	4.712	0.47	2.12
Paymaster 909	0.133	0.221	0.209	0.188	0.048	4.138	0.41	2.44
	0.142	0.191	0.174	0.169	0.025	3.685	0.37	2.70
	0.238	0.396	0.296	0.310	0.080	7.106	0.71	1.41
	0.145	0.147	0.154	0.149	0.005	3.191	0.32	3.13
CS7S	0.105	0.274	0.148	0.176	0.088	3.854	0.39	2.56
	0.147	0.164	0.139	0.150	0.013	3.223	0.32	3.13
	0.160	0.165	0.183	0.169	0.012	3.685	0.37	2.70
	0.177	0.232	0.210	0.206	0.028	4.582	0.46	2.17
Siokra 1-4	0.136	0.130	0.117	0.128	0.010	2.689	0.27	3.70
	0.222	0.215	0.220	0.219	0.004	4.898	0.49	2.04
	0.144	0.232	0.197	0.191	0.044	4.218	0.42	2.38
	0.179	0.227	0.134	0.180	0.047	3.951	0.40	2.50
Siokra L-22	0.140	0.137	0.154	0.144	0.009	3.078	0.31	3.23
	0.125	0.129	0.128	0.127	0.002	2.665	0.27	3.70
	0.116	0.151	0.168	0.145	0.027	3.102	0.31	3.23
	0.122	0.184	0.169	0.158	0.032	3.418	0.34	2.94
Strumica 5086	0.150	0.154	0.139	0.148	0.008	3.175	0.32	3.13
	0.168	0.160	0.162	0.163	0.004	3.539	0.35	2.86
	0.101	0.142	0.117	0.120	0.021	2.496	0.25	4.00
	0.136	0.132	0.156	0.141	0.013	3.005	0.30	3.33
Pima S-6	0.153	0.157	0.197	0.169	0.024	3.685	0.37	2.70
	0.110	0.147	0.166	0.141	0.028	3.005	0.30	3.33
	0.113	0.110	0.129	0.117	0.010	2.423	0.24	4.17
	0.074	0.155	0.137	0.122	0.043	2.544	0.25	4.00

APPENDIX 9

Calculation of number of cotton genomes in 1 μg of genomic DNA

Geever's (1980) estimation for the amount of DNA per cotton haploid genome

= 0.95 ρg of DNA/ haploid genome

$\Rightarrow$ 3.8 ρg of DNA/ tetraploid genome

So 1 μg of DNA = 1×10^6 ρg

= 2.63×10^5 tetraploid genomes

Alternatively:

1 ρg of DNA = 0.965×10^9 bp (Bennett and Leitch, 1995)

$\Rightarrow$ 3.667×10^9 bp/ tetraploid genome

And since 1 ρg of DNA = 0.965×10^9 bp

$\Rightarrow$ 1 μg of DNA = 0.965×10^{15} bp

$\Rightarrow$ 1 μg of DNA = 2.63×10^5 tetraploid genomes

APPENDIX 10

rRNA gene estimation: densitometer readings for Filter #1

a. Plasmid DNA standards

Amount of plasmid DNA (ηg)	Dilution (g/b)	Peak height (cm) data set #1	Peak height (cm) data set #2	Ave. peak height (cm)	St. dev.	Line of best fit for Filter #1
8	g	3.40	3.55	3.61	0.41	3.87
8	g	4.00	3.95			
8	b	2.75	3.60			
8	b	3.80	3.85			
10	g	3.10	4.20	4.03	0.50	4.46
10	g	4.25	4.50			
10	b	3.50	3.90			
10	b	4.25	4.55			
15	g	5.60	6.25	5.94	0.54	5.93
15	g	6.25	6.55			
15	b	4.90	5.60			
15	b	6.30	6.10			
20	g	7.30	7.20	7.95	0.74	7.40
20	g	7.35	7.45			
20	b	8.00	8.50			
20	b	8.70	9.10			
40	g	12.30	13.15	13.88	0.79	13.27
40	g	14.35	14.35			
40	b	14.05	14.55			
40	b	13.80	14.50			
60	g	16.10	16.75	18.66	2.68	19.15
60	g	18.50	19.60			
60	b	16.40	16.85			
60	b	22.10	23.00			

APPENDIX 10 (CONT'D)

rRNA gene estimation: densitometer readings for Filter #1

b. Genomic DNA samples, with variant values omitted

Cotton variety	Plant	Peak height (cm) data set #1	Peak height (cm) data set #2	Ave. peak height for plant (cm)*	Ave. peak height for variety (cm)*	St. dev. peak height for variety (cm)*
Paymaster 909	1	9.3	8.85	9.08	9.54	1.70
	2	10.6	11.95	11.28		
	4	7.4	8.15	7.86		
		7.5	8.4			
	5	9	9.7	10.59		
		10.95	12.7			
CS7S	1	6.55	7.25	7.58	7.31	0.62
		7.65	8.85			
	2	6.75	7	6.88		
	3	7.45	7.65	7.55		
	4	7.05	7.5	7.14		
		6.6	7.4			
Siokra 1-4	1	8.05	8.35	9.10	8.48	1.49
		9.3	10.7			
	4	7.35	7.4	9.01		
		10.1	11.2			
	5	7.4	7.6	7.34		
		6.55	7.8			
Siokra L-22	1	6.65	7.5	7.08	8.02	0.88
	2	7.05	6.95	8.00		
		8.4	9.6			
	3	8.25	8.45	8.35		
	4	8.55	8.6	8.35		
		7.5	8.75			
Strumica 5086	1	10.15	9.8	9.59	9.44	1.01
		8.6	9.8			
	2	8.25	8.35	9.56		
		10.4	11.25			
	5	8.25	8.75	9.16		
		9.2	10.45			
Pima S-6	1	9.65	10.15	10.59	10.54	1.60
		10.65	11.9			
	2	8.5	8.65	8.83		
		8.6	9.55			
	3	12.15	12.4	12.20		
		11.15	13.1			

* Calculation based on data sets #1 and #2 pooled

APPENDIX 11

rRNA gene estimation: densitometer readings for Filter #2

a. Plasmid DNA standards

Amount of plasmid DNA (ηg)	Dilution (r/b)	Peak height (cm) data set #1	Peak height (cm) data set #2	Ave. peak height (cm)	St. dev.	Line of best fit for Filter #2
8	r	2.9	2.95	2.96	0.30	3.69
8	r	2.7	2.5			
8	b	3.3	3.25			
8	b	3.3	2.8			
10	r	4.05	3.55	3.84	0.29	4.19
10	r	3.9	3.75			
10	b	3.55	3.5			
10	b	4.2	4.2			
15	r	6.55	6.45	5.81	0.65	5.45
15	r	5.15	4.65			
15	b	6	5.6			
15	b	6.1	6			
20	r	8	7.7	7.09	0.65	6.70
20	r	6.95	6.45			
20	b	6.3	6.45			
20	b	7.65	7.2			
40	r	14.15	13.7	12.90	1.12	11.71
40	r	13.9	13.7			
40	b	12.75	11.4			
40	b	12.15	11.45			
60	r	16.1	14.9	15.87	0.89	16.72
60	r	15.7	14.85			
60	b	16.2	16.2			
60	b	17.6	15.4			

APPENDIX 11 (CONT'D)

rRNA gene estimation: densitometer readings for Filter #2

b. Genomic DNA samples

Cotton variety	Plant	Peak height (cm) data set #1	Peak height (cm) data set #2	Ave. peak height for plant (cm)*	Ave. peak height for variety (cm)*	St. dev. peak height for variety (cm)*
Paymaster 909	1	7	6.6	6.80	7.40	0.54
	2	7.35	6.85	7.10		
	4	7.55	7.4	7.39		
		7.5	7.1			
	5	7.5	7.15	7.85		
		8.55	8.2			
CS7S	1	5.6	5.35	5.48	5.73	0.35
	2	5.75	5.6	5.64		
		5.9	5.3			
	3	5.95	5.95	6.09		
		6.45	6			
	4	5.6	5.25	5.43		
Siokra 1-4	1	7	7.1	7.46	6.73	0.88
		8	7.75			
	4	6.65	6.4	6.96		
		7.6	7.2			
	5	5.4	5.3	5.75		
		6.4	5.9			
Siokra L-22	2	6.4	6.2	6.35	6.18	0.34
		6.5	6.3			
	3	6.4	6.35	6.01		
		5.85	5.45			
	4	6.1	5.85	6.16		
		6.65	6.05			
Strumica 5086	1	6.7	6.55	7.06	7.25	0.60
		7.85	7.15			
	2	7.4	7.3	7.41		
		7.7	7.25			
	3	8.15	7.95	8.05		
	5	6.55	6.4	6.48		
Pima S-6	1	7.9	7.8	7.46	8.25	0.89
		7.3	6.85			
	2	7.8	7.7	8.08		
		8.65	8.15			
	3	9	8.5	8.75		
	4	9.85	9.55	9.70		

* Calculation based on data sets #1 and #2 pooled

APPENDIX 12

Statistical analysis of data for estimation of rRNA gene number

The data presented for analysis are shown below as : filter (1 or 2), species (1-6), animal-within-species, data#1, and data#2. The difference between data#1 and data#2 is also shown, and it is noted that these differences are almost invariably negative for filter 1 and positive for filter 2, The exceptions are indicated by $

1	1	1	9.30	8.85	0.45	$
1	1	2	10.60	11.95	-1.35	
1	1	3	7.40	8.15	-0.75	
1	1	4	7.50	8.40	-0.90	
1	1	5	9.00	9.70	-0.70	
1	1	6	10.95	12.70	-1.75	
1	2	21	6.55	7.25	-0.70	
1	2	22	7.65	8.85	-1.20	
1	2	23	6.75	7.00	-0.25	
1	2	24	7.45	7.65	-0.20	
1	2	25	7.05	7.50	-0.45	
1	2	26	6.60	7.40	-0.80	
1	3	31	8.05	8.35	-0.30	
1	3	32	9.30	10.70	-1.40	
1	3	33	7.35	7.40	-0.05	
1	3	34	10.10	11.20	-1.10	
1	3	35	7.40	7.60	-0.20	
1	3	36	6.55	7.80	-1.25	
1	4	41	6.65	7.50	-0.85	
1	4	42	7.05	6.95	0.10	$
1	4	43	8.40	9.60	-1.20	
1	4	44	8.25	8.45	-0.20	
1	4	45	8.55	8.60	-0.05	
1	4	46	7.50	8.75	-1.25	
1	5	51	10.15	9.80	0.35	$
1	5	52	8.60	9.80	-1.20	
1	5	53	8.25	8.35	-0.10	
1	5	54	10.40	11.25	-0.85	
1	5	55	8.25	8.75	-0.50	
1	5	56	9.20	10.45	-1.25	
1	6	61	9.65	10.15	-0.50	
1	6	62	10.65	11.90	-1.25	
1	6	63	8.50	8.65	-0.15	
1	6	64	8.60	9.55	-0.95	
1	6	65	12.15	12.40	-0.25	
1	6	66	11.15	13.10	-1.95	
2	1	71	7.00	6.60	0.40	
2	1	72	7.35	6.85	0.50	
2	1	73	7.55	7.40	0.15	
2	1	74	7.50	7.10	0.40	
2	1	75	7.50	7.15	0.35	
2	1	76	8.55	8.20	0.35	
2	2	81	5.60	5.35	0.25	
2	2	82	5.75	5.60	0.15	
2	2	83	5.90	5.30	0.60	
2	2	84	5.95	5.95	0.00	
2	2	85	6.45	6.00	0.45	
2	2	86	5.60	5.25	0.35	
2	3	91	7.00	7.10	-0.10	$
2	3	92	8.00	7.75	0.25	
2	3	93	6.65	6.40	0.25	
2	3	94	7.60	7.20	0.40	
2	3	95	5.40	5.30	0.10	
2	3	96	6.40	5.90	0.50	
2	4	101	6.40	6.20	0.20	

2	4	102	6.50	6.30	0.20
2	4	103	6.40	6.35	0.05
2	4	104	5.85	5.45	0.40
2	4	105	6.10	5.85	0.25
2	4	106	6.65	6.05	0.60
2	5	111	6.70	6.55	0.15
2	5	112	7.85	7.15	0.70
2	5	113	7.40	7.30	0.10
2	5	114	7.70	7.25	0.45
2	5	115	8.15	7.95	0.20
2	5	116	6.55	6.40	0.15
2	6	121	7.90	7.80	0.10
2	6	122	7.30	6.85	0.45
2	6	123	7.80	7.70	0.10
2	6	124	8.65	8.15	0.50
2	6	125	9.00	8.50	0.50
2	6	126	9.85	9.55	0.30

The analysis performed was a repeated measures analysis of variance with grouping factors of filter and species, and a within-animal factor of time. An unstructured error matrix was assumed. Program 5V from the BMDP Statistical Software Package (UCLA,1991) was used for analysis.

The results obtained for the significance or otherwise of the experimental effects are shown as :

WALD TESTS OF SIGNIFICANCE OF
FIXED EFFECTS AND COVARIATES

Effect	DF	CHI-SQUARE	P-VALUE
filter	1	84.64	0.000
species	5	78.67	0.000
time	1	16.00	0.000
filter*species	5	1.41	0.923
filter*time	1	101.16	0.000
species*time	5	0.99	0.963
filter*species*time	5	1.89	0.864

The estimate covariance and correlation matices for data#1 and data#2 are respectively :

.738	.796	1.00	.91
.796	1.036	.91	1.00

The predicted values for data#1 and data#2 for the various filter and species combinations are

F	S	E[data#1]	E[data#2]	
1	1	9.07319	9.76625	
1	2	7.19292	7.88598	
1	3	8.20519	8.89825	
1	4	7.77759	8.47065	
1	5	9.01792	9.71098	
1	6	9.98318	10.6762	i.e. approx .7 units increase d1->d2
2	1	7.60097	7.30236	
2	2	5.72070	5.42209	
2	3	6.73297	6.43436	
2	4	6.30537	6.00676	
2	5	7.54570	7.24709	
2	6	8.51096	8.21235	i.e. approx .3 units decrease d1->d2

These values are obtained from the fitted model using coefficients corresponding to the significant model effects of filter,species,time and filter*time.

APPENDIX 13

Independent calibration curves for rRNA gene estimation, derived from four sets of data

a. Filter #1, data set #1: $y = 0.28977x + 1.40463$

Amount of plasmid DNA (ηg)	Ave. peak height (cm)	St. dev.	Line of best fit
8	3.72	0.55	3.49
10	4.30	0.57	3.78
15	5.75	0.66	5.76
20	7.20	0.66	7.84
40	13.00	0.91	13.63
60	18.79	2.76	18.28

b. Filter #1, data set #2: $y = 0.29782x + 1.63891$

Amount of plasmid DNA (ηg)	Ave. peak height (cm)	St. dev.	Line of best fit
8	4.02	0.19	3.74
10	4.62	0.30	4.29
15	6.11	0.40	6.13
20	7.60	0.89	8.06
40	13.55	0.66	14.14
60	19.51	2.95	19.05

c. Filter #2, data set #1: $y = 0.25931x + 1.68540$

Amount of plasmid DNA (ηg)	Ave. peak height (cm)	St. dev.	Line of best fit
8	3.76	0.30	3.05
10	4.28	0.28	3.93
15	5.58	0.58	5.95
20	6.87	0.76	7.23
40	12.06	0.95	13.24
60	17.24	0.83	16.40

d. Filter #2, data set #2: $y = 0.24194x + 1.68886$

Amount of plasmid DNA (ηg)	Ave. peak height (cm)	St. dev.	Line of best fit
8	3.62	0.31	2.88
10	4.11	0.32	3.75
15	5.32	0.77	5.68
20	6.53	0.61	6.95
40	11.37	1.31	12.56
60	16.21	0.63	15.34

REFERENCES

Aguilar F, Montandon P-E, Stutz E (1991) Two genes encoding the soybean translation elongation factor eEF-1α are transcribed in seedling leaves. *Plant Mol Biol* **17**: 351-360

Alexander DC, Williams B, Delmer DP (1983) Gene expression in the developing cotton fibre. *Plant Physiol* **72(S)**: 147

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410

Amor Y, Haigler C, Johnson S, Wainscott M, Delmer DP (1995) A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. *Proc Natl Acad Sci USA* **92**: 9353-9357

Appels R, Moran LB, Gustafson JP (1986) The structure of DNA from the rye (*Secale cereale*) NOR R1 locus and its behaviour in wheat backgrounds. *Can J Genet Cytol* **28**: 673-685

Arondel V, Kader J-C (1990) Lipid transfer in plants. *Experentia* **46**: 579-585

Augspurger CK (1986) Morphology and dispersal potential of wind-dispersed diospores of neo-tropical trees. *Amer J Bot* **73**: 353-363

Ausio J (1992) Structure and dynamics of transcriptionally active chromatin. J Cell Sci **102**: 1-5

Ausubel F M, Brent R, Kingston RE, Moore DD, Seidmann JG, Smith JA, Struhl K (1987) *Current Protocols in Molecular Biology*. John Wiley and Sons Inc, NY

Averyhart-Fullard V, Datta K, Marcus A (1988) A hydroxyproline-rich protein in the soybean cell wall. *Proc Natl Acad Sci USA* **85**: 1082-1085

Axelos M, Bardet C, Liboz T, Le Van Thai A, Curie C, Lescure B (1989) The gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1α: molecular cloning, characterisation and expression. *Mol Gen Genet* **219**: 106-112

Bacic A, Harris PJ, Stone BA (1988) Structure and function of plant cell walls. In Stumpf PK, Conn EE (eds) *The Biochemistry of Plants. A Comprehensive Treatise* **14**: 297-371. Academic Press, Inc

Bailey NTJ (1959) The correlation of measurements. In *Statistical Methods in Biology*. The English Universities Press Ltd, pp 78-90

Baker RJ, Longmire JL, Van Den Busshe RA (1995) Organisation of repetitive elements in the Upland cotton genome (*Gossypium hirsutum* L.). *J Hered* **86**: 178-185

Baker RV, Griffin AC (1984) Ginning. *In* Kohel RJ, Lewis CF (eds) *Cotton.* American Society of Agronomy Inc, pp 397-436

Balch WE, Kahn RA, Schwaninger R (1992) ADP-ribosylation factor is required for vesicular trafficking between the endoplasmic reticulum and the *cis*-Golgi compartment. *J Biol Chem* **267**: 13053-13061

Baluška F, Parker JS, Barlow PW (1993) A role for gibberellic acid in orienting microtubules and regulating cell growth polarity in the maize root cortex. *Planta* **191**: 149-157

Basra AS (1982) Metabolic regulation of fibre elongation in a short- and long-staple cotton. PhD Dissertation, Punjab Agricultural University, Ludhiana, India

Basra AS, Malik CP (1981) *In* Verma SC (ed) *Contemporary Trends in Plant Sciences.* Kalyani, New Delhi, pp 490-496

Basra AS, Malik CP (1983) Dark metabolism of CO_2 during fibre elongation of two cottons differing in fibre lengths. *J Exp Bot* **34**: 1-9

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

Basra AS, Sarlach RS, Nayyar H, Malik CP (1990) Sucrose hydrolysis in relation to development of cotton (*Gossypium* spp.) fibres. *Ind J Exp Bot* **28**: 985-988

Beasley CA (1973) Hormonal regulation of growth in unfertilised cotton ovules. *Science* **179**: 1003-1005

Beasley CA (1977) Temperature-dependent response to indoleacatic acid is altered by NH_4^+ in cultured cotton ovules. *Plant Physiol* **59**: 203-206

Beasley CA (1979) Cellulose content in fibres of cottons which differ in their lint lengths and extent of fuzz. *Physiol Plant* **45**: 77-82

Beasley CA, Ting IP (1973) The effects of plant growth substances on *in vitro* fibre development from fertilised cotton ovules. *Amer J Bot* **60**: 130-139

Beasley CA, Ting IP (1974) Effects of plant growth substances on *in vitro* fibre development from unfertilised cotton ovules. *Amer J Bot* **61**: 188-194

Benedict CR, Smith RH, Kohel RJ (1973) Incorporation of ^{14}C-photosynthate into developing cotton bolls, *Gossypium hirsutum* L. *Crop Sci* **13**: 98-91

Benedict CR, Kohel RJ, Jividen GM (1994) Crystalline cellulose and cotton fibre strength. *Crop Sci* **34**: 147-151

Bennett MD, Leitch IJ (1995) Nuclear DNA amounts in angiosperms. *Ann Bot* **76**: 113-176

Berberich T, Sugawara K, Harada M, Kusano T (1995) Molecular cloning, characterisation and expression of an elongation factor 1α gene in maize. *Plant Mol Biol* **29**: 611-615

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

Betrabet SM, Pillay KPR, Iyengar RLN (1963) Structural properties of cotton fibres: Part II. Birefringence and structural reversals in relation to mechanical properties. *Text Res J* **33**: 720-727

Binder S, Thalheim C, Brennicke A (1994) Transcription of potato mitochondrial 26S rRNA is initiated at its mature 5' end. *Curr Genet* **26**: 519-523

Birnstiel ML, Chipchase M, Spiers J (1971) The ribosomal RNA cistrons. *Prog Nucl Acid Res Mol Biol* **11**: 351-389

Bonen L, Boer PH, McIntosh JE, Gray MW (1987) Nucleotide sequence of the wheat mitochondrial gene for subunit I of cytochrome oxidase. *Nucl Acids Res* **15**: 6734

Boylston EK, Hebert JJ, Hensarling TP, Bradow JM, Thibodeaux DP (1990) Role of silicon in developing cotton fibres. *J Plant Nut* **13**: 131-148

Bradley DJ, Kjellbom P, Lamb CJ (1992) Elicitor- and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: a novel, rapid defense response. *Cell* **70**: 21-30

Brands JHGM, Maasen JA, van Hemert FJ, Amons R, Möller W (1986) The primary structure of the α subunit of human elongation factor 1. *Eur J Biochem* **155**: 167-171

Brandstädter J, Rossbach C, Theres K (1994) The pattern of histone H4 expression in the tomato shoot apex changes during development. *Planta* **192**: 69-74

Brown JWS (1986) A catalogue of splice junction and putative branch point sequences from plant introns. *Nucl Acids Res* **14**: 9549-9559

Brown RM, Li L, Okuda K, Kuga S, Kudlicka K, Drake R, Santos R, Clement S (1994) *In vitro* synthesis of cellulose in plants. *Plant Physiol* **105**: 1-2

Buescher PJ, Phillips RL, Brambl R (1984) Ribosomal RNA contents of maize genotrophs with different ribosomal RNA gene numbers. *Biochem Genet* **22**: 923-930

Bullock WO, Fernandez JM, Short JM (1987) XL1-Blue: A high efficiency plasmid transforming *rec*A *Escherichia coli* strain with beta-galactosidase selection. *Biotech* **5**: 376-378

Bulone V, Fincher GB, Stone BA (1995) *In vitro* synthesis of a microfibrillar (1→3)-β-glucan by a ryegrass (*Lolium multiflorum*) endosperm (1→3)-β-glucan synthase enriched by product entrapment. *Plant J* **8**: 213-225

Burton K (1956) A study of the conditions and mechanism of the diphenylamine reaction for the colourimetric estimation of deoxyribonucleic acid. *Biochem* **62**: 315-323

Busslinger M, Portman R, Irminger JC, Birnstiel ML (1980) Ubiquitous and gene-specific regulatory 5' sequences in a sea urchin histone DNA clone coding for histone protein variants. *Nucl Acids Res* **8**: 957-977

Cammue BPA, Thevissen K, Hendriks M, Eggermont K, Goderis IJ, Proost P, Van Damme J, Osborne RW, Guerbette F, Kader J-C, Broekaert (1995) A potent antimicrobial protein from onion seed showing sequence homology to plant lipid transfer proteins. *Plant Physiol* **109**: 445-455

Carpita NC (1984) Cell wall development in maize coleoptiles. *Plant Physiol* **76**: 205-212

Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J* **3**: 1-30

Cassab GI, Varner JE (1988) Cell wall proteins. *Ann Rev Plant Physiol Plant Mol Biol* **39**: 321-353

Chaboute M-E, Chaubet N, Phillips G, Ehling M, Gigot C (1987) Genomic organisation and nucleotide sequence of two histone H3 and two histone H4 genes of *Arabidopsis thaliana*. *Plant Mol Biol* **8**: 179-191

Chen J, Varner JE (1985) Isolation and characterisation of cDNA clones for carrot extensin and a proline-rich 33-kDa protein. *Proc Natl Acad Sci USA* **82**: 4399-4403

Cherry JP, Leffler HP (1984) Seed. *In* Kohel RJ, Lewis CF (eds) *Cotton*. American Society of Agronomy Inc, pp 511-569

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochem* **18**: 5294-5299

Clarke L, Carbon J (1976) A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *E coli* genome. *Cell* **9**: 91-99

Colton RT (1991) Cotton fibre quality. *Agfacts*, New South Wales Dept Ag

Condeelis J (1995) Elongation factor 1α, translation and the cytoskeleton. *Trends Biochem Sci* **20**:169-170

Conner JW, Krieg DR, Gipson JR (1972) Accumulation of simple sugars in developing cotton bolls as influenced by night temperatures. *Crop Sci* **12**: 752-754

Cooper JB, Chen JA, van Holst G-J, Varner, JE (1987) Hydroxyproline-rich glycoproteins of plant cell walls. *Trends Biochem Sci* **12**: 24-27

Corell RA, Myler P, Stuart K (1994) *Trypanosoma brucei* mitochondrial CR4 gene encodes an extensively edited mRNA with completely edited sequence only in bloodstream forms. *Mol Biochem Parasitol* **64**: 65-74

Cosgrove DJ (1993) How do plant cell walls extend? *Plant Physiol* **102**: 1-6

Cosgrove DJ, Durachko DM (1994) Autolysis and extension of isolated walls from growing cucumber hypocotyls. *J Exp Bot* **45**: 1711-1719

Cottrelle P, Cool M, Thuriaux, Price V, Theile D, Buhler J, Fromageot P (1985) Either one of two yeast EF-1A genes is required for cell viability. *Curr Genet* **9**: 693-697

Cousins YL, Lyon BR, Llewellyn DJ (1991) Transformation of an Australian cotton cultivar: prospects for cotton improvement through genetic engineering. *Aust J Plant Physiol* **18**: 481-494

Cox RA (1968) The use of guanidinium chloride in the isolation of nucleic acids. *Methods Enzymol* **12B**: 120-129

Crane CF, Price HJ, Stelly DM, Czeschin, DG, McKnight TD (1993) Identification of a homeologous chromosome pair by *in situ* DNA hybridisation to ribosomal RNA loci in meiotic chromosomes of cotton (*Gossypium hirsutum*). *Genome* **36**: 1015-1022

Crouse EJ, Schmitt JM, Bohnert H-J, Gordon k, Driesel AJ, Herrmann RG (1978) Intramolecular compositional heterogeneity of *Spinacia* and *Euglena* chloroplast DNAs. *In* Akoyunoglou G, Argyroudi-Akoyunoglou JH (eds) *Chloroplast Development.* Elsevier/North-Holland, NY, pp 565-572

Cullis CA (1990) DNA rearrangements in response to environmental stress. *Adv Genet* **28**: 73-97

Curie C, Liboz T, Bardet C, Gander E, Médale C, Axelos M, Lescure B (1991) *Cis* and *trans*-acting elements involve in the activation of *Arabidopsis thaliana* A1 gene encoding the translation elongation factor EF-1α. *Nucl Acids Res* **19**: 1305-1310

Curie C, Liboz T, Montané M-H, Rouan D, Axelos M, Lescure B (1992) The activation process of *Arabidopsis thaliana* A1 gene encoding the translation elongation factor EF-1α is conserved among angiosperms. *Plant Mol Biol* **18**: 1083-1089

Curie C, Axelos M, Bardet C, Atanassova R, Chaubet N, Lescure B (1993) Modular organisation and developmental activity of an *Arabidopsis thaliana* EF-1α gene promoter. *Mol Gen Genet* **238**: 428-436

Dale RM, Mendu N, Ginsburg H, Kridl JC (1984) Sequence analysis of the maize mitochondrial 26S rRNA gene and flanking regions. *Plasmid* **11**: 141-150

Datta K, Marcus A (1990) Nucleotide sequence of a gene encoding soybean repetitive proline-rich protein 3. *Plant Mol Biol* **14**: 285-286

Datta K, Schmidt A, Marcus A (1989) Characterisation of two soybean repetitive proline-rich proteins and a cognate cDNA from germinated axes. *Plant Cell* **1**: 945-952

Davidonis GH (1989) Fibre development in preanthesis cotton ovules. *Physiol Plant* **17**: 290-294

Davidonis G, Hinojosa O (1994) Influence of seed location on cotton fibre development in planta and *in vitro*. *Plant Sci* **203**: 107-113

de Silva J, Arrowsmith D, Hellyer A, Whiteman S, Robinson S (1994) Xyloglucan transferase and plant growth. *J Exp Bot* **45**: 1693-1701

DeLanghe E, Kosmidou-Dimitropoulou S, Waterkeyn L (1978) Effect of hormones on nucleolar growth and vacuolation in elongating cotton fibres. *Planta* **140**: 269-273

Dellaporta SL, Wood J, Hicks JB (1985) Maize DNA miniprep. *In Molecular Biology of Plants: A Laboratory Course Manual.* Cold Spring Harbour Laboratory, NY, pp 36-37

Delmer DP (1987) Cellulose biosynthesis. *Ann Rev Plant Physiol* **38**: 259-290

Delmer DP, Solomon M, Read SM (1991) Direct photolabeling with [^{32}P]UDP-glucose for identification of a subunit of cotton fibre callose synthase. *Plant Physiol* **95**: 556-563

Delmer DP, Ohana P, Gonen L, Benziman M (1993) *In vitro* synthesis of cellulose in plants: still a long way to go! *Plant Physiol* **103**: 307-308

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

Deltour R, Brouchart R (1971) Root cell ultrastructure of *Zea mays* embryo during early stages of germination. *Planta* **97**: 197-207

Denecke J, De Rycke R, Botterman J (1992) Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope. *EMBO J* **11**: 2345-2355

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl Acids Res* **12**: 387-395

Dhindsa RS, Beasley CA, Ting IP (1975) Osmoregulation in cotton fibre: accumulation of potassium and malate during growth. *Plant Physiol* **56**: 394-398

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

Djé MK, Mazabraud A, Viel A, le Maire M, Denis H, Crawford E, Brown DD (1990) Three genes under different developmental control encode elongation factor 1α in *Xenopus laevis*. *Nucl Acids Res* **18**: 3489-3493

Doerfler W (1983) DNA methylation and gene activity. *Ann Rev Biochem* **52**: 93-124

Domingo C, Gómez MD, Cañas L, Hernández-yago J, Conejero V, Vera P (1994) A novel, extracellular matrix protein from tomato associated with lignified secondary walls. *Plant Cell* **6**: 1035-1047

Duggleby RG, Kinns H, Rood JI (1981) A computer program for determining the size of DNA restriction fragments. *Anal Biochem* **110**:49-55

Dunn MA, Morris A, Jack PL, Hughes MA (1993) A low temperature-responsive trnslation factor 1α from barley (*Hordeum vulgare* L.). *Plant Mol Biol* **23**: 221-225

Durso NA, Cyr RJ (1994a) A calmodulin-sensitive interaction between microtubules and a higher plant homologue of elongation factor-1α. *Plant Cell* **6**: 893-905

Durso NA, Cyr RJ (1994b) Beyond translation: elongation factor-1α (EF-1α) and the cytoskeleton. *Protoplasma* **180**: 99-105

Dynan WS (1986) Promoters for housekeeping genes. *Trends Genet* **2**: 196-197

Ebener W, Fowler TJ, Suzuki H, Shaver J, Tierney ML (1993) Expression of DcPRP1 is linked to carrot storage root formation and is induced by wounding and auxin treatment. *Plant Physiol* **101**: 259-265

Edwards GA, Endrizzi JE, Stein R (1974) Genome DNA evolution and chromosome organisation in Gossypium. *Chromosoma* **47**: 309-326

Edwards JW, Coruzzi GM (1990) Cell-specific gene expression in plants. *Ann Rev Genet* **24**: 275-303

Efstratiadis A, Posakony A, Maniatis T, Lawn R, O'Connell C, Spiritz R, DeRiel J, Forget B, Weissman S, Slightom J, Blechl A, Smithies O, Baralle F, Shoulders C, Proudfoot N (1980) The structure and evolution of the human β-globin gene family. *Cell* **21**: 653-668

Endrizzi JE, Turcotte EL, Kohel RJ (1984) Qualitative genetics, cytology, and cytogenetics. *In* Kohel RJ, Lewis CF (eds) *Cotton*. American Society of Agronomy Inc, pp 81-129

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

Falconet D, Sevignac M, Quetier F (1988) Nucleotide sequence and determination of the extremeties of the 26S ribosomal RNA gene in wheat mitochondria: evidence for sequence rearrangements in the ribosomal genes of higher plants. *Curr Genet* **13**: 75-82

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction fragments to high specific activity. *Anal Biochem* **132**: 6-13

Finer JJ, McMullen MD (1990) Transformation of cotton (*Gossypium hirsutum* L.) via particle bombardment. *Plant Cell Rep* **8**: 586-589

Firoozabady E (1986) The effects of cell cycle parameters on cell wall regeneration and cell division of cotton protoplasts (*Gossypium hirsutum* L.). *J Exp Bot* **37**: 1211-1217

Firoozabady E, DeBoer DL, Merlo DJ, Halk EL, Amerson LN, Rashka KE, Murray EE (1987) Transformation of cotton (*Gossypium hirsutum* L.) by *Agrobacterium tumefaciens* and regeneration of transgenic plants. *Plant Mol Biol* **10**: 105-116

Flavell RB, O'Dell M, Thompson WF (1988) Regulation of cytosine methylation in ribosomal DNA and nucleolus organiser expression in wheat. *J Mol Biol* **204**: 523-534

Flavell RB, O'Dell M, Vincentz M, Sardana R, Barker RF (1986) The differential expression of ribosomal RNA genes. *Philos Trans R Soc Lond B* **314**: 385-397

Flavell RB, O'Dell M, Smith DB, Thompson WF (1985) Chromosome architecture: the distribution of recombination sites, the structure of ribosomal DNA loci and the multiplicity of sequences containing inverted repeats. *In* Van Vloten-Doting L, Groot GSP, Hall TC (eds) *Molecular Form and Function of the Plant Genome*. Plenum Press, NY, pp 1-14

Fleming AJ, Mandel T, Hofmann S, Sterk P, de Vries SC, Kuhlemeier C (1992) Expression pattern of a tobacco lipid transfer protein gene within the shoot apex. *Plant J* **2**: 855-862

Flint EA (1950) The structure and development of the cotton fibre. *Biol Rev* **25**: 414-434

Forde BG (1994) AT-rich elements (ATREs) in the promoter regions of nodulin and other higher plant genes: a novel class of *cis*-acting regulatory element? *In* L Nover (ed) *Results and problems in cell differentiation* **20**: 87-103. Springer-Verlag, Heidelburg

Frischauf A-M, Lehrach H, Poustka A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. *J Mol Biol* **170**: 827-842

Fry SC (1986) Cross-linking of matrix polymers in growing cell walls of angiosperms. *Ann Rev Plant Physiol* **37**: 165-186

Fry SC (1994) Unzipped by expansins. *Curr Biol* **4**: 815-817

Fryxell PA (1979) *The natural history of the cotton tribe (Malvaceae, Tribe Gossypieae).* Texas A&M University Press, College Station, TX

Galau GA, Bass HW, Hughes DW (1988) Restriction fragment length polymorphisms in diploid and allotetraploid *Gossypium*: Assigning the late embryogenesis-abundant (*Lea*) alloalleles in *G. hirsutum*. *Mol Gen Genet* **211**: 305-314

Galau GA, Legocki AB, Greenway SC, Dure LS (1981) Cotton messenger RNA sequences exist in both polyadenylated and nonpolyadenylated forms. *J Biol Chem* **256**: 2551-2560

Gallie DR, Kobayashi M (1994) The role of the 3'-untranslated region of non-polyadenylated plant viral mRNAs in regulating translational efficiency. *Gene* **142**: 159-165

Gayler KR, Glasziou KT (1972) Physiological functions of acid and neutral invertases in growth and sugar storage in sugarcane. *Physiol Plant* **27**: 25-31

Geever RF (1980) The evolution of single-copy nucleotide sequences in the genomes of *Gossypium hirsutum* L. PhD Dissertation, University of Arizona, Tucson

Ghosh S (1987) The nucleolus. *Int Rev Cytol* **17**: 573-598

Gipson JR (1986) Temperature effects on growth, development, and fibre properties. In Mauney JR, Stewart JMcD (eds), *Cotton Physiology*. The Cotton Foundation, Memphis, TN, pp 47-56

Gipson JR, Joham HE (1969) Influence of night temperature on growth and development of cotton (*Gossypium hirsutum* L.) III. Fibre elongation. *Crop Sci* **9**: 127-129

Gipson JR, Ray LL (1969) Fibre elongation rates in five varieties of cotton (*Gossypium hirsutum* L.) as influenced by night temperature. *Crop Sci* **9**: 339-341

Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. *Nat Genet* **3**: 266-272

Goodall GJ, Filipowicz W (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* **58**: 473-483

Goodall GJ, Filipowicz W (1991) Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. *EMBO J* **10**: 2635-2644

Graves DA, Stewart J McD (1988a) Analysis of the protein constituency of developing cotton fibres. *J Exp Bot* **39**: 59-69

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

Gray MW, Huh TY, Schnare MN, Spencer DF, Falconet D (1983) Organisation and evolution of ribosomal RNA genes in wheat mitochondria. *In* Cifeni O, Dure L (eds) *Structure and Function of Plant Genomes*. Plenum Press, NY, pp 373-380

Grimes DW, Miller RJ, Wiley PL (1975) Cotton and corn development in two field soils of different characteristics. *Agron J* **67**: 519-523

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. *Nature* **292**: 21-26

Grunstein M, Hogness D (1975) Colony hybridisation: a method for the isolation of cloned DNAs that contain a specific gene. *Proc Natl Acad Sci USA* **72**: 3961

Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. *Gene* **25**: 263-269

Guilfoyle TJ, Lin CY, Chen YM, Nagao RT, Key JL (1975) Enhancement of soybean RNA polymerase I by auxin. *Proc Natl Acad Sci USA* **72**: 69-72

Haaf T, Hayman DL, Schmid M (1991) Quantitative determination of rDNA transcription units in vertebrate cells. *Exp Cell Res* **193**: 78-86

Haigler CH, Rao NR, Roberts EM, Huang J-Y, Upchurch DR, Trolinder NL (1991) Cultured ovules as models for cotton fibre development under low temperatures. *Plant Physiol* **95**: 88-96

Hammond D (1941) The expresssion of genes for leaf shape in *Gossypium hirsutum* L. and *Gossypium arboreum* L. I. The expresssion of genes for leaf shape in *Gossypium hirsutum* L. *Am J Bot* **28**: 124-138

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**: 557-580

Hasenfratz M-P, Tsou C-L, Wilkins TA (1995) Expression of two related vacoular H+-ATPase 16-kilodalton proteolipid genes is differentially regulated in a tissue-specific manner. *Plant Physiol* **108**: 1395-1404

Hawkins RS, Serviss GH (1930) Development of cotton fibres in the Pima and Acala varieties. *J Ag Res* **40**: 1017-1029

Hay B, Short JM (1992) ExAssist™ helper phage and SOLR™ cells for Lambda ZAP®II excisions. *Strategies* **5**: 16-18

Hayashi T, Read SM, Bussell J, Thelen M, Lin F-C, Brown RM, Delmer DP (1987) UDP-glucose: (1→3)-β-Glucan synthases from mung bean and cotton. *Plant Physiol* **83**: 1054-1062

He M, Wilde A, Kaderbhai MA (1990) A simple single-step procedure for small-scale preparation of *E. coli* plasmids. *Nucl Acids Res* **18**: 1660

Heidecker G, Messing J (1986) Structural analysis of plant genes. *Ann Rev Plant Physiol* **37**: 439-466

Hemleben V, Grierson D (1978) Evidence that in higher plants the 25S and 18S genes are not interspersed with genes for 5S rRNA. *Chromosoma* **65**: 353-358

Hemleben V, Zentgraf U (1994) Structural organisation and regulation of transcription by RNA polymerase I of plant nuclear ribosomal RNA genes. In Nover L (ed) *Results and Problems in Cell Differentiation* **20**: 3-24. Springer-Verlag, Heidelburg

Herman P, Marks MD (1989) Trichome development in Arabidopsis thaliana. II. Isolation and complementation of the *GLABROUS1* gene. *Plant Cell* **1**: 1051-1055

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237-244

Higo H, Kishimoto N, Saito A, Higo K (1994) Molecular cloning and characterisation of a cDNA encoding a small GTP-binding protein related to mammalian ADP-ribosylation factor from rice. *Plant Sci* **100**: 41-49

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

Hodge R, Paul W, Draper J, Scott R (1992) Cold-plaque screening: a simple technique for the isolation of low abundance, differentially expressed transcripts from conventional cDNA libraries. *Plant J* **2**: 257-260

Holmes DS, Quigley M (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* **114**: 193-197

Hong JC, Nagao, RT, Key JL (1987) Characterisation and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean. *J Biol Chem* **262**: 8367-8376

Hong JC, Nagao, RT, Key JL (1990) Characterisation of a proline-rich cell wall protein gene family of soybean. *J Biol Chem* **265**: 2470-2475

Horsfall JG (1972) *Genetic Vulnerability of Major Crops*. National Academy of Sciences, Washington DC, pp 269-281

Hovemann B, Richter S, Walldorf U, Cziepluch C (1988) Two genes encode related cytoplasmic elongation factors 1α (EF-1α) in *Drosophila melanogaster* with continuous and stage-specific expression. *Nucl Acids Res* **16**: 3175-3194

Hughes MA, Dunn MA, Pearce RS, White AJ, Zhang L (1992) An abscisic-acid-responsive, low temperature barley gene has homology with a maize phospholipid transfer protein. *Plant Cell Env* **15**: 861-865

Hülskamp M, Miséra S, Jürgens G (1994) Genetic dissection of trichome cell development in *Arabidopsis*. *Cell* **76**: 555-566

Hunt AG (1994) Messenger RNA 3' end formation in plants. *Ann Rev Plant Physiol Plant Mol Biol* **45**: 47-60

Hutchinson J (1959) *The Application of Genetics to Cotton Improvement*. Cambridge University Press, Cambridge

Huwyler HR, Franz G, Meier H (1979) Changes in the composition of cotton fibre cell walls during development. *Planta* **146**: 635-642

Huynh TV, Young RA, Davis RW (1985) Constructing and screening cDNA libraries in λgt10 and λgt11. *In* Glover DM (ed) *DNA Cloning: A Practical Approach* **1**: 49-78. IRL Press, Oxford

Ingle J (1963) The extraction and estimation of nucleotides and nucleic acids from plant material. *Phytochem* **2**: 353-370

Ingle J, Timmis JN, Sinclair J (1975) The relationship between satellite DNA, ribosomal RNA gene redundancy, and genome size in plants. *Plant Physiol* **55**: 496-501

Inoue T, Zhong HS, Miyao A, Ashikawa I, Monna L, Fukuoka S, Miyadera N, Nagamura Y, Kurata N, Sasaki T, Minobe Y (1994) Sequence-tagged sites (STSs) as standard landmarkers in the rice genome. *Theor Appl Genet* **89**: 728-734

Isaac PG, Jones VP, Leaver CJ (1985) The maize cytochrome *c* oxidase subunit I gene: sequence, expression and rearrangement in cytoplasmic male sterile plants. *EMBO J* **4**: 1617-1623

Jaquet J-P, Buchala AJ, Meier H (1982) Changes in the non-structural carbohydrate content of cotton (*Gossypium* spp.) fibres at different stages of development. *Planta* **156**: 481-486

Jaswandala RT, Singh YD, Chinoy JJ (1977) Auxin metabolism in developing cotton hairs. *J Exp Bot* **28**: 1111-1116

John ME (1992) An efficient method for isolation of RNA and DNA from plants containing polyphenolics. *Nucl Acids Res* **20**: 2381

John ME (1994) Re-engineering cotton fibre. *Chem Ind* **17**: 676-679

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, Stewart JMcD (1992) Genes for genes: biotechnological advances in cotton. *Trends Biotech* **10**: 165-170

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

Johnson JM (1969) A study of nucleolar vacuoles in cultured tobacco cells using radioautography, actinomycin D, and electron microscopy. *J Cell Biol* **43**: 197-206

Josè M, Puigdomènech P (1993) Structure and expression of genes coding for structural proteins of the plant cell wall. *New Phytol* **125**: 259-282

Joshi CP (1987a) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. *Nucl Acids Res* **15**: 9627-9640

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

Joshi PA, Stewart JMcD, Graham ET (1988) Ultrastructural localisation of ATPase activity in cotton fibre development during elongation. *Protoplasma* **143**: 1-10

Jupe SC, Scott IM (1989) The *slender* phenotype of pea: stem growth, peroxidase levels and ethylene responses. *Physiol Plant* **77**: 59-66

Jupe SC, Scott IM (1992) Gibberellin and the *pro* gene suppress peroxidase activity in elongating tomato (*Lycopersicon esculentum* Mill.) stem tissues. *Ann Bot* **69**: 33-37

Kader J-C (1990) Intracellular transfer of phospholipids, galactolipids, and fatty acids in plant cells. *Subcell Biochem* **16**: 69-111

Kadir ZBA (1976) DNA evolution in the genus *Gossypium*. *Chromosoma* **56**: 85-94

Kadowaki K, Suzuki T, Kazama S, Oh-fuchi T, Sakamoto W (1989) Nucleotide sequence of the cytochrome oxidase subunit I gene from rice mitochondria. *Nucl Acids Res* **17**: 7519

Kalla R, Shimamoto K, Potter R, Stein Nielson P, Linnestad C, Olsen O-A (1994) The promoter of the barley aleurone-specific gene encoding a putative 7 kDa lipid transfer protein confers aleurone cell-specific expression in transgenic rice. *Plant J* **6**: 849-860

Karn J, Matthes WD, gait MJ, Brenner S (1984) A new selective phage cloning vector, λ2001, with sites for *Xba*I, *Bam*HI, *Hin*dIII, *Eco*RI, *Sst*I and *Xho*I. *Gene* **32**: 217-224

Kasperbauer MJ (1994) Cotton plant size and fibre developmental responses to FR/R ratio reflected from the soil surface. *Physiol Plant* **91**: 317-321

Katterman FRH, Shattuck, VI (1983) An effective method of DNA isolation from the mature leaves of *Gossypium* species that contain large amounts of phenolic terpenoids and tannins. *Prep Biochem* **13**: 347-359

Kavanagh T.A., Timmis J.N. (1986). Heterogeneity in cucumber ribosomal DNA. *Theor. Appl. Genet.* **72**: 337-345.

Kawahara R, Sunabori S, Fukuda H, Komamine A (1992) A gene expressed preferentially in the globular stage of somatic embryogenesis encodes elongation-factor 1α in carrot. *Eur J Biochem* **209**: 157-162

Kawahara R, Sunabori S, Fukuda H, Komamine A (1994) Analysis by *in situ* hybridisation of the expression of elongation factor 1α in the carrot cells during somatic embryogenesis. *J Plant Res* **107**: 361-364

Kearns EV, Keck P, Somerville CR (1992) Primary structure of cytochrome b_5 from cauliflower (*Brassica oleracea* L.) deduced from peptide and cDNA sequences. *Plant Physiol* **99**: 1254-1257

Keller B (1993) Structural cell wall proteins. *Plant Physiol* **101**: 1127-1130

Keller B, Lamb CJ (1989) Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation. *Gene Devel* **3**: 1639-1646

Keller B, Sauer N, Lamb CJ (1988) Glycine-rich cell wall proteins in bean: gene structure and association of the protein with the vascular system. *EMBO J* **7**: 3625-3633

Keller W (1995) No end yet to messenger RNA 3' processing! *Cell* **81**: 829-832

Kim DW, Uetsuki T, Kaziro Y, Yamaguchi N, Sugano S (1990) Use of the human elongation factor 1α promoter as a versatile and efficient expression system. *Gene* **91**: 217-223

Kiss T, Kis M, Solymosy F (1989) Nucleotide sequence of a 25S rRNA gene from tomato. *Nucl Acids Res* 17: 796

Kiyosue T, Shinozaki K (1995) Cloning of a carrot cDNA for a member of the family of ADP-ribosylation factors (ARFs) and characterisation of the binding of nucleotides by its product after expression in *E. coli. Plant Cell Physiol* **36**: 849-856

Kleis-San Francisco SM, Tierney ML (1990) Isolation and characterisation of a proline-rich cell wall protein from soybean seedling. *Plant Physiol* **94**: 1897-1902

Kloth RH (1989) Changes in the level of tubulin subunits during development of cotton (*Gossypium hirsutum*) fibre. *Physiol Plant* **76**: 37-41

Kloth RH (1992) Variability of malate dehydrogenase among cotton cultivars with differing fibre traits. *Crop Sci* **32**: 617-621

Kohel RJ, Quisenberry JE, Benedict CR (1974) Fibre elongation and dry weight changes in mutant lines of cotton. *Crop Sci* **14**: 471-474

Kohel RJ, Narbuth EV, Benedict CR (1992) Fibre development of Ligon lintless-2 mutant of cotton. *Crop Sci* **32**: 733-735

Kohel RJ, Benedict CR, Jividen GM (1993) Incorporation of [^{14}C]glucose into crystalline cellulose in aberrant fibres of a cotton mutant. *Crop Sci* **33**: 1036-1040

Kolattukudy PE (1981) Structure, biosynthesis, and biodegradation of cutin and suberin. *Ann Rev Plant Physiol* **32**: 539-567

Kolosha VO, Fodor I (1990) Nucleotide sequence of *Citrus limon* 26S rRNA gene and secondary structure model of its RNA. *Plant Mol Biol* **14**: 147-161

Kolosha VO, Kryukov VM, Fodor I (1986) Sequence analysis of *Citrus limon* DNA coding for 26S rRNA. *FEBS Lett* **197**: 89-92

Koltunow AM, Truettner J, Cox KH, Wallroth M, Goldberg RB (1990) Different temporal and spatial gene expression patterns occur during anther development. *Plant Cell* **2**: 1201-1224

Kotilainen M, Helariutta Y, Elomaa P, Paulin L, Teeri TH (1994) A corolla- and carpel-abundant, non-specific lipid transfer protein gene is expressed in the epidermis and parenchyma of *Gerbera hybrida* var. Regina (Compositae). *Plant Mol Biol* **26**: 971-978.

Kozak M (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucl Acids Res* **12**: 857-872

Kudlicka K, Brown RM, Li L, Lee JH, Shin H, Kuga S (1995) β-glucan synthesis in the cotton fibre. IV. *In vitro* assembly of the cellulose I allomorph. *Plant Physiol* **107**: 111-123

Kurnick NB (1952) Histological staining with methyl green-pyronin. *Stain Tech* **27**: 233-242

Kyte J, Doolittle, RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**: 105-132

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680

Lamport DA (1965) The protein component of primary cell walls. *Adv Bot Res* **2**: 151-218

Lang AG (1938) The origin of lint and fuzz hairs of cotton. *J Agric Res* **56**: 507-521

Langer RHM, Hill GD (1982) Malvaceae. *In Agricultural Plants*, 1st edn. Cambridge Univ Press, pp 259-263

Lawrence GJ, Appels R (1986) Mapping the nucleolus organiser region, seed protein loci and isozyme loci on chromosome 1R in rye. *Theor Appl Genet* **71**: 742-749

Leaver CJ, Ingle J (1971) The molecular integrity of chloroplast ribosomal ribonucleic acid. *Biochem J* **123**: 235-243

Lee JA (1984) Cotton as a world crop. *In* Kohel RJ, Lewis CF (eds) *Cotton.* American Society of Agronomy Inc, pp 1-25

Li L, Brown RM (1993) β-glucan synthesis in the cotton fibre. II. Regulation and kinetic properties of β-glucan synthases. *Plant Physiol* **101**: 1143-1148

Li Q, Hunt AG (1995) A near-upstream element in a plant polyadenylation signal consists of more than six nucleotides. *Plant Mol Biol* **28**: 927-934

Li L, Drake RR, Clement S, Brown RM (1993) β-glucan synthesis in the cotton fibre. III. Identification of UDP-glucose-binding subunits of β-glucan synthases by photoaffinity labelling with [β-^{32}P]5'-N_3-UDP-glucose. *Plant Physiol* **101**: 1149-1156

Lin FC, Brown RM, Drake RR, Haley BE (1990) Identification of the uridine-5'-diphosphoglucose (UDP-glc) binding subunit of cellulose synthase in *Acetobacter xylinum* using the photoaffinity probe 5'-azido-UDP-glucose. *J Biol Chem* **265**: 4782-4784

Linnestad C, Lönneborg A, Kalla R, Olsen O-A (1991) Promoter of a lipid transfer protein gene expressed in barley aleurone cells contains similar *myb* and *myc* recognition sites as the maize *Bz-McC* allele. *Plant Physiol* **97**: 841-843

Löbler M, Hirsch AM (1993) A gene that encodes a proline-rich nodulin with limited homology to PsENOD12 is expressed in the invasion zone of *Rhizobium meliloti*-induced alfalfa root nodules. *Plant Physiol* **103**: 21-30

Luehrsen KR, Walbot V (1994) Intron creation and polyadenylation in maize are directed by AU-rich RNA. *Genes Devel* **8**: 1117-1130

Lütcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG codon differs in plants and animals. *EMBO J* **6**: 43-48

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

MacNeish RS (1964) Ancient mesoamerican civilisation. *Science* **143**: 531-537

Madsen HO, Poulsen K, Dahl O, Clark BFC, Hjorth JP (1990) Retropseudogenes constitute the major part of the human elongation factor 1α gene family. *Nucl Acids Res* **18**: 1513-1516

Mahe A, Grisvard J, Dron M (1992) Fungal- and plant-specific gene markers to follow the bean anthracnose infection process and normalise a bean chitinase mRNA induction. *Mol Plant-Microbe Interact* **5**: 242-248

Maltby D, Carpita NC, Montezinos D, Kulow C, Delmer DP (1979) β-1,3-glucan in developing cotton fibres. *Plant Physiol* **63**: 1158-1164

Marcus A, Greenberg J, Averyhart-Fullard V (1991) Repetitive proline-rich proteins in the extracellular matrix of the plant cell. *Physiol Plant* **81**: 273-279

Mayer R, Ross P, Weinhouse H, Amikam D, Volman G, Ohana P, Calhoon RD, Wong HC, Emerick AW, Benziman M (1991) Polypeptide composition of bacterial cyclic diguanylic acid-dependent cellulose synthase and the occurrence of immunologically crossreacting proteins in higher plants. *Proc Natl Acad Sci USA* **88**: 5472-5476

McCabe DE, Martinell BJ (1993) Transformation of elite cotton cultivars via particle bombardment of meristems. *Biotech* **11**: 596-598

McCann MC, Roberts K (1994) Changes in cell wall architecture during cell elongation. *J Exp Bot* **45**: 1683-1691

McCann MC, Wells B, Roberts K (1990) Direct visualisation of cross-links in the primary plant cell wall. *J Cell Sci* **96**: 323-334

McClure BA, Hagen G, Brown CS, Gee MA, Guilfoyle TJ (1989) Transcription, organisation and sequence of an auxin-regulated gene cluster in soybean. *Plant Cell* **1**: 229-239

McDonell MW, Simon MN, Studier FW (1977) Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *J Mol Biol* **110**: 119-146

McKnight SL, Miller OL (1976) Ultrastructural patterns of RNA synthesis during early embryogenesis of *Drosophila melanogaster*. *Cell* **8**: 305-319

McQueen-Mason SJ (1995) Expansins and cell wall expansion. *J Exp Bot* **46**: 1639-1650

McQueen-Mason SJ, Cosgrove DJ (1994) Disruption of hydrogen bonding between plant cell wall polymers by proteins that induce wall extension. *Proc Natl Acad Sci USA* **91**: 6574-6578

McQueen-Mason SJ, Cosgrove DJ (1995) Expansin mode of action on cell walls. *Plant Physiol* **107**: 87-100

McQueen-Mason S, Durachko DM, Cosgrove DJ (1992) Two endogenous proteins that induce cell wall extension in plants. *Plant Cell* **4**: 1452-1433

Meier H, Buchs L, Buchala AJ, Homewood T (1981) (1→3)-β-D-Glucan (callose) is a probable intermediate in biosynthesis of cellulose of cotton fibres. *Nature* **289**: 821-822

Meijer EA, de Vries SC, Sterk P, Gadella DWJ, Wirtz KWA, Hendriks T (1993) Characterisation of the non-specific lipid transfer protein EP2 from carrot (*Daucus carota* L.). *Mol Cell Biochem* **123**: 159-166

Meinert MC, Delmer DP (1977) Changes in biochemical composition of the cell wall of the cotton fibre during development. *Plant Physiol* **59**: 1088-1097

Memon AR, Kawazoe R, Zhang X, Herrin DL, Thompson GA (1995) Low molecular mass GTP-binding proteins in *Chlamydomonas reinhardtii* wilt type and a wall-less strain. Characterisation and comparison with GTP-binding proteins of *Dunaliella salina*. *Plant Physiol Biochem* **33**: 225-234

Meredith WR (1984) Quantitative genetics. *In* Kohel RJ, Lewis CF (eds) *Cotton*. American Society of Agronomy Inc, pp 131-150

Meredith WR, Pettigrew WT, Heitholt JJ (1996) Sub-okra, semi-smoothness, and nectariless effect on cotton lint yield. *Crop Sci* **36**: 22-25

Meredith WR, Culp TW, Robert KQ, Ruppenicker GF, Anthony WS, Williford JR (1991) Determining future cotton variety fibre quality objectives. *Textile Res J* **61**: 715-720

Metz AM, Timmer RT, Allen ML, Browning KS (1992) Sequence of a cDNA encoding the α-subunit of wheat translation elongation factor 1. *Gene* **120**: 315-316

Michaelson MJ, Price HJ, Ellison JR, Johnston JS (1991) Comparison of plant DNA contents determined by feulgen microspectrophotometry and laser flow cytometry. *Am J Bot* **78**: 183-188

Miller PA, Rawlings JO (1967) Breakup of initial linkage blocks through intermating in cotton breeding populations. *Crop Sci* **7**: 199-204

Mizusawa S, Nishimura S, Seela F (1986) Improvement of the dideoxy chain termination method of DNA sequencing by the use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucl Acids Res* **14**: 1319-1324

Mogen BD, MacDonald MH, Leggewie G, Hunt AG (1992) Several distinct types of sequence elements are required for efficient mRNA 3' end formation in pea *rbcS* gene. *Mol Cell Biol* **12**: 5406-5414

Molina A, García-Olmedo F (1993) Developmental and pathogen-induced expression of three barley genes encoding lipid transfer proteins. *Plant J* **4**: 983-991

Moore PJ, Swords KMM, Lynch MA, Staehelin LA (1991) Spatial organisation of the assembly pathway of glycoproteins and complex polysaccharides in the Golgi apparatus of plants. *J Cell Biol* **112**: 589-602

Morelli JK, Shewmaker CK, Vayda ME (1994) Biphasic stimulation of translational activity correlates with induction of EF-1α upon wounding in potato tubers. *Plant Physiol* **106**: 897-903

Mundy J, Rogers JC (1986) Selective expression of a probable amylase/protease inhibitor in barley aleurone cells: comparison to the barley amylase/subtilisin inhibitor. *Planta* **169**: 51-63

Murphy G, Kavanagh T (1988) Speeding-up the sequencing of double-stranded DNA. *Nucl Acids Res* **16**:5198

Murray NJ, Williamson NP, Lilley TH, Haslam E (1994) Study of the interaction between salivary proline-rich proteins and a polyphenol by 1H-NMR spectroscopy. *Eur J Biochem* **219**: 923-935

Nagata S, Nagashima K, Tsunetsugu-Yokota Y, Fujimura K, Miyazaki M, Kaziro Y (1984) Polypeptide chain elongation factor 1α (EF-1α) from yeast: Nucleotide sequence of one of the two genes for EF-1α from *Saccharomyces cerevisiae*. *EMBO J* **3**: 1825-1830

Naithani SC, Rama Rao N, Singh YD (1982) Physiological and biochemical changes associated with cotton fibre development. *Physiol Plant* **54**: 225-229

Naithani SC, Rama Rao N, Krishnan PN, Singh YD (1981) Changes in o-diphenol oxidase during fibre development in cotton. *Ann Bot* **48**: 379-385

Nelson M, McClelland M (1987) The effect of site-specific methylation on restriction-modification enzymes. *Nucl Acids Res* **15**: 219-230

Neumann GM, Condron R, Thomas I, Polya GM (1994) Purification and sequencing of a family of wheat lipid transfer protein homologues phosphorylated by plant calcium-dependent protein kinase. *Biochim Biophys Acta* **1209**: 183-190

Newman T, de Bruijn FD, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E, Somerville C (1994) Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol* **106**: 1241-1255

Ni Ghogain N, Byrne H, Timmis J (1982) The genetic control of ribosomal RNA accumulation in flax genotrophs. Heredity **48**: 211-226

Nickrent DL, Soltis DE (1995) A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbc*L sequences. *Ann Missouri Bot Gard* **82**: 208-234

Nishitani K, Tominaga R (1992) Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyses transfer of a segment of xyloglucan molecule to another xyloglucan molecule. *J Biol Chem* **267**: 21058-21064

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

Odell JT, Knowlton S, Lin W, Mauvis J (1988) Properties of an isolated transcription stimulating sequence derived from the cauliflower mosaic virus 35S promoter. *Plant Mol Biol* **10**: 263-272

O'Kelley JC (1953) The use of C^{14} in locating growth regions in the cell walls of elongating cotton fibres. *Plant Physiol* **28**: 281-286

O'Kelley JC, Carr PH (1953) Elongation of the cotton fibre. *In* Loomis WE (ed) *Growth and Differentiation in Plants.* Iowa Stata University Press, Ames, Iowa, pp 55-68

Okuda K, Li L, Kudlicka K, Kuga S, Brown RM (1993) β-glucan synthesis in the cotton fibre. I. Identification of β-1,4- and β-1,3-glucans synthesised *in vitro*. *Plant Physiol* **101**: 1131-1142

O'Neill SD, Nadeau JA, Zhang XS, Bui AQ, Halevy AH (1993) Interorgan regulation of ethylene biosynthetic genes by pollination. *Plant Cell* **5**: 419-432

Østergaard J, Højrup P, Knudsen J (1995) Amino acid sequences of three acyl-binding/lipid-transfer proteins from rape seedlings. *Biochim Biophys Acta* **1254**: 169-179

Østergaard J, Vergnolle C, Schoentgen F, Kader J-C (1993) Acyl-binding/lipid-transfer proteins from rape seedlings, a novel ctagory of proteins interacting with lipids. *Biochim Biophys Acta* **1170**: 109-117

Ozols J (1989) Structure of cytochrome b_5 and its topology in the microsomal membrane. *Biochim Biophys Acta* **997**: 121-130

Palmer JD (1985) Comparative organization of chloroplast genomes. *Ann Rev Genet* **19**: 325-354

Palmer JD, Thompson WF (1981) Clone banks of the mung bean, pea and spinach chloroplast genomes. *Gene* **15**: 21-26

Peeters M-C, Dillemans W, Voets (1991) Nucleolar activity in differentiating cotton fibres is related to the position of the boll on the plant. *J Exp Bot* **42**: 353-357

Peeters M-C, Voets S, Dayatilake G, DeLanghe E (1987) Nucleolar size at early stages of cotton fibre development in relation to final fibre dimension. *Physiol Plant* **71**: 436-440

Peeters M-C, Voets S, Wijsmans J, DeLanghe E (1988) Pattern of nucleolar growth in differentiating cotton fibres (*Gossypium hirsutum* L.). *Ann Bot* **62**: 377-382

Pelèse-Siebenbourg F, Caelles C, Kader J-C, Delseny M, Puigdomènech P (1994) A pair of genes coding for lipid-transfer proteins in *Sorghum vulgare*. *Gene* **148**: 305-308

Percy RG, Turcotte EL (1992) Interspecific hybrid fibre characteristics of cotton altered by unconventional *Gossypium barbadense* L. fibre genotypes. *Crop Sci* **32**: 1437-1441

Perkins HH, Ethridge DE, Bragg CK (1984) Fibre. *In* Kohel RJ, Lewis CF (eds) *Cotton.* American Society of Agronomy Inc, pp 437-509

Perlak FJ, Deaton RW, Armstrong TA, Fuchs RL, Sims SR, Greenplate JT, Fischhoff (1990) Insect resistant cotton plants. *Biotech* **8**: 939-943

Plant AL, Cohen A, Moses MS, Bray EA (1991) Nucleotide sequence and spatial expression pattern of a drought- and abscisic acid-induced gene of tomato. *Plant Physiol* **97**: 900-906

Pokalsky AR, Hiatt WR, Ridge N, Rasmussen R, Houck CM, Shewmaker CK (1989) Structure and expression of elongation factor 1α in tomato. *Nucl Acids Res* **17**: 4661-4673

Porter DD (1936) Positions of seeds and motes in locks and lengths of cotton fibres from bolls borne at different positions on plants at Greenville, TX. *USDA Tech Bull* **509**

Potikha T, Delmer DP (1995) A mutant of *Arabidopsis thaliana* displaying altered patterns of cellulose deposition. *Plant J* **7**: 453-460

Preston RD (1979) Polysaccharide conformation and cell wall function. *Ann Rev Plant Physiol* **30**: 55-78

Pyee J, Kolattukudy PE (1995) The gene for the major cuticular wax-associated protein and three homologous genes from broccoli (*Brassica oleracea*) and their expression patterns. *Plant J* **7**: 49-59

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

Qi G, Cao G, Jiang P, Feng X, Gu X (1988) Studies on the sites expressing evolutionary changes in the structure of eukaryotic 5S ribosomal RNA. *J Mol Evol* **27**: 336-340

Qi X, Behrens BX, West PR, Mort AJ (1995) Solubilisation and partial characterisation of extensin fragments from cell walls of cotton suspension cultures. *Plant Physiol* **108**: 1691-1701

Ramsey JC, Berlin, JD (1976a) Ultrastructure of early stages of cotton fibre elongation. *Am J Bot* **63**: 868-876

Ramsey JC, Berlin, JD (1976b) Ultrastructure of early stages of cotton fibre differentiation. *Bot Gaz* **137**: 11-19

Raven PH, Evert RF, Eichorn SE (1992) Biology of Plants. 4th edn, Worth Publishers, NY

Reeder RH (1990) rRNA synthesis in the nucleolus. *Trends Genet* **6**: 390-395

Reeder RH, Roan JG, Dunaway M (1983) Spacer regulation of *Xenopus* ribosomal gene transcription: competition in oocytes. *Cell* **35**: 449-456

Regad F, Bardet C, Tremousaygue D, Moisan A, Lescure B, Axelos M (1993) cDNA cloning and expression of an *Arabidopsis* GTP-binding protein of the ARF family. FEBS **316**: 133-136

Reid PE, Thomson NJ (1993) Development of *Verticillum* wilt tolerant cotton for Australia. *In* Imrie BC, Hacker JB (eds) *Proceedings of the 10th Australian Plant Breeding Conference*, Gold Coast, Australia, pp 149-150

Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organisation and evolution in a disomoic polyploid genome. *Genetics* **138**: 829-847

Rich PR, Bendall DS (1975) Cytochrome components of plant microsomes. *Eur J Biochem* **55**: 333-341

Riis B, Rattan SIS, Clark BFC, Merrick, WC (1990) Eukaryotic protein elongation factors. *Trends Biochem Sci* **15**: 420-424

Risueño MC, Medina FJ (1986) The nucleolar structure in plant cells. *Cell Biol Rev* **7**: 1-140

Risueño MC, Testillano PS (1994) Cytochemistry and immunocytochemistry of nucleolar chromatin in plants. *Micron* **25**: 331-360

Risueño MC, Testillano PS, Sanchez-Pina MA (1988) Variations of nucleolar ultrastructure in relation to transcriptional activity during G1, S and G2 periods of microspore interphase. *In* Cresti M, Gori P, Paccini E (eds) *Sexual Reproduction in Higher Plants*. Springer-Verlag, Berlin, Heidelberg, pp 9-14

Robert LS, Allard S, Gerster JL, Cass L, Simmonds J (1994) Molecular cloning of two *Brassica napus* genes expressed in the stigma. *Plant Mol Biol* **26**: 1217-1222

Robles J, Doers M (1994) pGEM®-T vector systems troubleshooting guide. *Promega Notes* **45**: 19-20

Rogers SO, Bendich AJ (1987a) Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Mol Biol* **9**: 509-520

Rogers SO, Bendich AJ (1987b) Heritability and variability in ribosomal RNA genes of *Vicia faba*. *Genetics* **117**: 285-295

Rose RJ, Setterfield G, Fowke LC (1972) Activation of nucleoli in tuber slices and the function of nucleolar vacuoles. *Exp Cell Res* **71**: 1-16

Roth BA, Goff SA, Klein TM, Fromm ME (1991) *C1*- and *R*-dependent expression of the maize *Bz1* gene requires sequences with homologies to mammalian myb and myc binding sites. *Plant Cell* **3**: 317-325

Rothnie HM, Reid J, Hohn T (1994) The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3'-end formation in plants. *EMBO J* **13**: 2200-2210

Ryser U (1977) Cell wall growth in elongating cotton fibres. *Cytobiologie* **15**: 78-84

Ryser U (1985) Cell wall biosynthesis in differentiating cotton fibres. *Eur J Cell Biol* **39**: 236-256

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: A laboratory manual*, 2nd edn. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N.Y.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463-5467

Sardana R, O'Dell M, Flavell R (1993) Correlation between the size of the intergenic regulatory region, the status of cytosine methylation of rRNA genes and nucleolar expression in wheat *Mol Gene Genet* **236**: 155-162

Sargent TD (1987) Isolation of differentially expressed genes. *Methods Enzymol* **152**: 423-432

Sasaki T, Song J, Koga-Ban Y, Matsui E, Fang F, Higo H, Nagasaki H, Hori M, Miya M, Murayama-Kayano E, Takiguchi T, Takasuga A, Niki T, Ishimaru K, Ikeda H, Yamamoto Y, Mukai Y, Ohta I, Miyadera N, Havukkala I, Minobe Y (1994) Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *Plant J* **6**: 615-624

Sato S, Myoraku A (1994) Three-dimensional organisation of nucleolar DNA in the higher plant nucleonolema studied by electron microscopy. *Micron* **25**: 431-437

Sawadogo M, Van Dyke MW (1991) A rapid method for the purification of deprotected oligodeoxynucleotides. *Nucl Acids Res* **19**: 674

Saxena R, Lin FC, Brown M (1991) Identification of a new gene in an operon for cellulose biosynthesis in *Acetobacter xylinum*. *Plant Mol Biol* **16**: 947-954

Saxena IM, Kudlicka K, Okuda K, Brown RM (1994) Characterisation of genes in the cellulose-synthesising operon (*acs* operon) of *Acetobacter xylinum*: implications for cellulose crystallisation. *J Bacter* **176**: 5735-5752

Schmidt JS, Lindstrom JT, Vodkin LO (1994) Genetic length polymorphisms create size variation in proline-rich proteins of the cell wall. *Plant J* **6**: 177-186

Schubert AM, Benedict CR, Berlin JD, Kohel RJ (1973) Cotton fibre development-kinetics of cell elongation and secondary wall thickening. *Crop Sci* **13**: 704-709

Schünmann PHD, Harrison J, Ougham HJ (1994) *Slender* barley, an extension growth mutant. *J Exp Bot* **45**: 1753-1760

Scott NS, Possingham JV (1980) Chloroplast DNA in expanding spinach leaves. *J Exp Bot* **123**: 1081-1092

Scott NS, Kavanagh TA, Timmis JN (1984) Methylation of rRNA genes in some higher plants. *Plant Sci Lett* **35**: 213-217

Seagull RW (1986) Changes in microtubule organisation and wall microfibril orientation during *in vitro* cotton fibre development: an immunofluorescent study. *Can J Bot* **64**: 1373-1381

Seagull RW (1992) A quantitative electron microscopic study of changes in microtubule arrays and wall microfibril orientation during *in vitro* cotton fibre development. *J Cell Sci* **101**: 561-577

Seagull RW (1993) Cytoskeletal involvement in cotton fibre growth and development. *Micron* **24**: 643-660

Seed B, Parker RC, Davidson N (1982) Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. *Gene* **19**: 201-209

Segura A, Moreno M, García-Olmedo F (1993) Purification and antipathogenic activity of lipid transfer proteins (LTPs) from the leaves of *Arabidopsis* and spinach. *FEBS Lett* **3**: 243-246

Senda M, Harada T, Mikami T, Sugiura M, Kinoshita T (1991) Genomic organisation and sequence analysis of the cytochrome oxidase subunit II gene from normal and male-sterile mitochondria in suger beet. *Curr Genet* **19**: 175-181

Shaw PJ, Highett MI, Beven AF, Jordan EG (1995) The nucleolar architecture of polymerase I transcription and processing. *EMBO J* **14**: 2896-2906

Shcherban TY, Shi J, Durachko DM, Guiltinan MJ, McQueen-Mason SJ, Shieh M, Cosgrove DJ (1995) Molecular cloning and sequence analysis of expansins- a highly conserved, multigene family of proteins that mediate cell wall extension in plants. *Proc Natl Acad Sci USA* **92**: 9245-9249

Sheng J, D'Ovidio R, Mehdy MC (1991) Negative and positive regulation of a novel proline-rich mRNA by fungal elicitor and wounding. *Plant J* **1**: 345-354

Shenk T (1981) Transcriptional control regions: nucleotide sequence requirements for initiation by RNA Polymerase II and III. *Curr Top Microbiol Immunol* 93: **25**-46

Shewmaker CK, Ridge NP, Pokalsky AR, Rose RE, Hiatt WR (1990) Nucleotide sequence of an EF-1α genomic clone from tomato. *Nucl Acids Res* **18**: 4276

Shin DH, Lee JY, Hwang KY, Kim KK, Suh SW (1995) High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure* **3**: 189-199

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete sequence of the tobacco chloroplast genome: its organization and expression. *EMBO J* **5**: 2043-2049.

Short JM, Fernandez JM, Sorge JA, Huse WD (1988) λZAP: a bacteriophage λ expression vector with *in vivo* excision properties. *Nucl Acids Res* **16**: 7583-7600

Showalter AM (1993) Structure and function of plant cell wall proteins. *Plant Cell* **5**: 9-23

Showalter AM, Varner JE (1989) Plant hydroxyproline-rich glycoproteins. *In* Stumpf PK, Conn EE (eds), *The Biochemistry of Plants* **15**: 485-420. Academic Press, NY

Showalter AM, Rumeau D (1990) Molecular biology of the plant cell hydroxyproline-rich glycoproteins. *In* RP Mecham, WS Adair (eds) *Organisation and Assembly of Plant and Animal Extracellular Matrix*. Academic Press, NY, pp 247-281

Silflow CD, Oppenheimer DG, Kopczak SD, Ploense SE, Ludwig SR, Haas N, Snustad DP (1987) Plant tubulin genes: structure and differential expression during development. *Dev Genet* **8**: 435-460

Skriver K, Leah R, Müller-Uri F, Olsen F-L, Mundy J (1992) Structure and expression of the barley lipid transfer protein gene *Ltp1*. *Plant Mol Biol* **18**: 585-589

Slobin LI (1980) The role of eukaryotic elongation factor Tu in protein synthesis. *Eur J Biochem* **110**: 555-563

Smillie RM, Scott NS (1969). Organelle biosynthesis: the chloroplast. In: *Progress in molecular biology*, vol.I. Springer, Berlin, Hedielberg, NY, pp 136-202

Smith MA, Stobart AK, Shewry PR, Napier JA (1994) Tobacco cytochrome b_5: cDNA isolation, expression analysis and in vitro protein targeting. *Plant Mol Biol* **25**: 527-537

Sollner-Webb B, Tower J (1986) Transcription of cloned eukaryotic ribosomal RNA genes. *Ann Rev Biochem* **55**: 801-830

Sollner-Webb B, Mougey EB (1991) News from the nucleolus: rRNA gene expression. *Trends Biochem Sci* **16**: 58-62

Staden R (1980) A new computer method for the storage and manipulation of DNA gel reading data. *Nucl Acids Res* **8**: 3673-3694

Staden R (1984) Computer methods to aid the determination and analysis of DNA sequences. *Biochem Soc Trans* **12**: 1005-1008

Standal R, Iversen T-G, Coucheron DH, Fjærvik E, Blatny JM, Valla S (1994) A new gene required for cellulose production and a gene encoding cellulolytic activity in *Acetobacter xylinum* are colocalised with the *bcs* operon. *J Bacteriol* **176**: 665-672

Steiner S, Philippsen P (1994) Sequence and promoter analysis of the highly expressed *TEF* gene of the filamentous fungus *Ashbya gossypii*. *Mol Gen Genet* **242**: 263-271

Stephens SG (1944) Phenogenetic evidence for the amphidiploid origin of New World cottons. *Nature* **153**: 53-54

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

Stewart J McD (1975) Fibre initiation on the cotton ovule (*Gossypium hirsutum*). *Amer J Bot* **62(7)**: 723-730

Stewart SD, Sterling WL (1989) Causes and temporal patterns of cotton fruit abscission. *J Econ Entomol* **82**: 954-959

Sturkie DG (1934) A study of lint and seed development in cotton as influenced by environmental factors. *Am Soc Agron* **26**: 1-24

Sugiura M, Iida Y, Oono K, Takaiwa F (1985) The complete nucleotide sequence of a rice 25S rRNA gene. *Gene* **37**: 255-259

Svaren J, Hörz W (1993) Histones. Nucleosomes and transcription. *Curr Opin Genet Dev* **3**: 219-225

Szopa J, Müller-Röber B (1994) Cloning and sequence analysis of an ADP-ribosylation factor from *Solanum tuberosum* L. *Plant Cell Rep* **14**: 180-183

Taft EB (1951) The problem of a standardised technic for the methyl green-pyronin stain. *Stain Tech* **26**: 205-212

Talbott LD, Ray PM (1992) Molecular size and separability features of pea cell wall polysaccharides: implications for models of primary wall structure. *Plant Physiol* **98**: 357-368

Tchang F, This P, Stiefel V, Arondel V, Morch M-D, Pages M, Puigdomenech P, Grellet F, Delseny M, Bouillon P, Huet J-C, Guerbette F, Beauvais-Cante F, Duranton H, Pernollett J-C, Kader J-C (1988) Phospholipid transfer protein: full-length cDNA and amino acid sequence in maize. *J Biol Chem* **263**: 16849-16855

Thaker VS, Saroop S, Vaishnav PP, Singh YD (1989) Genotypic variations and influence of diurnal temperature on cotton fibre development. *Field Crops Res* **22**: 1-13

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

Thompson DM, Tanzer MM, Meagher RB (1992) Degradation products of the mRNA encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in soybean and transgenic petunia. *Plant Cell* **4**: 47-58

Thomson NJ (1971) Heterosis and combining ability of American and African cotton cultivars in a low latitude under high-yield conditions. *Aust J Agric Res* **22**: 759-770

Thomson NJ (1973a) Intra-varietal variability and response to single plant selection in *Gossypium hirsutum* L. 1. Phenotypic variability. *J Agric Sci Camb* **80**: 135-145

Thomson NJ (1973b) Intra-varietal variability and response to single plant selection in *Gossypium hirsutum* L. 2. Genotypic variability *J Agric Sci Camb* **80**: 147-160

Thomson NJ (1973c) Intra-varietal variability and response to single plant selection in *Gossypium hirsutum* L. 3. Response to selection *J Agric Sci Camb* **80**: 161-170

Thomson NJ (1979) Cotton. In Lovett JV, Lazenby A (eds) *Australian Field Crops* **2**: 113-136. Angus and Robertson Publishers, Aust.

Thomson NJ, Luckette DJ (1988) Heterosis and combining ability effects in cotton. II Heterosis. *Aust J Agric Res* **39**: 991-1002

Thomson NJ, Reid PE, Williams ER (1987) Effects of the okra leaf, nectariless, frego bract and glabrous conditions on yield and quality of cotton lines. *Euphytica* **36**: 545-553

Thuring RWJ, Sanders JPM, Borst P (1975) A freeze-squeeze method for recovering long DNA from agarose gels. *Anal Biochem* **66**: 213-220

Timmis JN, Ingle J (1973) Environmentally induced changes in rRNA gene redundancy. *Nature* **244**: 235-236

Timmis JN, Ingle J (1975) Quantitative regulation of gene activity in plants. *Plant Physiol* **56**: 255-258

Timmis JN, Sinclair J, Ingle J (1972) Ribosomal RNA genes in euploids and aneuploids of hyacinth. *Cell Diff* **1**: 335-339

Timpa JD, Ramey HH (1989) Molecular characterisation of three cotton varieties. *Text Res J* **59**: 661-664

Timpa JD, Triplett BA (1993) Analysis of cell wall polymers during cotton fibre development. *Planta* **189**: 101-108

Tiwari SC, Wilkins TA (1995) Cotton (*Gossypium hirsutum*) seed trichomes expand via diffuse growing mechanism. *Can J Bot* **73**: 746-757

Torres RA, Ganal M, Hemleben V (1990) GC balance in the internal transcribed spacers ITS1 and ITS2 of nuclear ribosomal RNA genes. *J Mol Evol* **30**: 170-181

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

Triplett JD, Timpa BA (1995) Characterisation of cell wall polymers from cotton ovule culture fibre cells by gel permeation chromatography. *In Vitro Cell Dev Biol- Plant* **31**: 171-175

Trolinder NL, Goodin JR (1987) Somatic embryogenesis and plant regeneration in cotton (*Gossypium hirsutum* L.). *Plant Cell Rep* **6**: 231-234

Tsuboi S, Suga T, Takishima K, Mamiya G, Matsui K, Ozeki Y, Yamada M (1991) Organ-specific occurrence and expression of the isoforms of nonspecific lipid transfer protein in castor bean seedlings, and molecular cloning of a full-length cDNA for a cotyledon-specific isoform. *J Biochem* **110**: 823-831

Umbeck P, Johnson G, Barton K, Swain W (1987) Genetically transformed cotton (*Gossypium hirsutum* L.) plants. *Biotech* **5**: 263-266

Ursin VM, Irvine JM, Hiatt WR, Shewmaker CK (1991) Developmental analysis of elongation factor EF-1α expression in transgenic tobacco. *Plant Cell* **3**: 583-591

van der Pijl L (1982) *Principles of dispersal in Higher Plants*. 3rd edn. Springer-Verlag, Berlin

Varner JE, Lin L-S (1989) Plant cell wall architecture. *Cell* **56**: 231-239

Vayda ME, Shewmaker CK, Morelli JK (1995) Translational arrest in hypoxic potato tubers is correlated with the aberrant association of elongation factor EF-1α with polysomes. *Plant Mol Biol* **28**: 751-757

Vedel F, Delseny M (1987) Repetitivity and variability of higher plant genomes. *Plant Physiol Biochem* **25**: 191-210

Verma DPS, Cheon C-I, Hong Z (1994) Small GTP-binding proteins and membrane biogenesis in plants. *Plant Physiol* **106**: 1-6

Verwoert IIGS, Brown A, Slabas AR, Stuitje AR (1995) A *Zea mays* GTP-binding protein of the ARF family complements an *Escherichia coli* mutant with a temperature-sensitive malonyl-coenzyme A:acyl carrier protein transacylase. *Plant Mol Biol* **27**: 629-633

Vignols F, Lund G, Pammi S, Trémousaygue D, Grellet F, Kader J-C, Puigdomènech P, Delseny M (1994) Characterisation of a rice gene coding for a lipid transfer protein. *Gene* **142**: 265-270

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

von Heijne G (1986) A new method for predicting signal sequence cleavage sites. *Nucl Acids Res* **14**: 4683-4690

Walbot V, Cullis CA (1985) Rapid genomic change in higher plants. *Ann Rev Plant Physiol* **36**: 367-396

Walbot V, Dure LS (1976) Developmental biochemistry of cotton seed embryogenesis and germination. VII Characterisation of the cotton genome. *J Mol Biol* **101**: 503-536

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

Ward OP, Moo-Young M (1989) Enzymatic degradation of cell wall and related plant polsaccharides. *CRC Crit Rev Biotech* **8**: 237-274

Waterkeyn L (1981) Cytochemical localisation and function of the 3-linked glucan callose in the developing cotton fibre cell wall. *Protoplasma* **106**: 49-67

Webster GC (1985) Protein synthesis in aging organisms. *In* Sohal RS (ed) *Molecular Biology of Aging: Gene Stability and Gene Expression*. Raven Press, NY, pp 263-289

Wendel JF (1989) New World tetraploid cottons contain Old World cytoplasm. *Proc Natl Acad Sci USA* **86**: 4132-4136

Wendel JF, Schnabel A, Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc Natl Acad Sci USA* **92**: 280-284

Westafer JM, Brown RM (1976) Electron microscopy of the cotton fibre: new observations on cell wall formation. *Cytobios* **15**: 111-138

Whitney SEC, Brigham JE, Darke AH, Reid JSG, Gidley MJ (1995) *In vitro* assembly of cellulose/xyloglucan networks: ultrastructural and molecular aspects. *Plant J* **8**: 491-504

Wiebauer K, Herrero J-J, Filipowicz W (1988) Nuclear pre-mRNA processing in plants: distinct modes of 3'-splice-site selection in plants and animals. *Mol Cell Biol* **8**: 2042-2051

Wilkins TA, Wan C-Y (1993) Members of two vacuolar H^+-ATPase subunit A gene families are spatially and temporally regulated in developing trichomes. *Plant Physiol* **102** (S): 10

Williamson MP (1994) The structure and function of proline-rich regions in proteins. *Biochem J.* **297**: 249-260

Willison JHM, Brown RM (1977) An examination of the developing cotton fibre: wall and plasmalemma. *Protoplasma* **92**: 21-41

Willmitzer L (1988) The use of transgenic plants to study plant gene expression. *Trends Genet* **4**: 13-18

Wilson RC, Cooper JB (1994) Characterisation of PRP1 and PRP2 from *Medicago trunculata*. *Plant Physiol* **105**: 445-446

Wilson RC, Long F, Maruoka EM, Cooper, JB (1994) A new proline-rich early nodulin from *Medicago trunculata* is highly expressed in nodule meristematic cells. *Plant Cell* **6**: 1265-1275

Wolffe AP (1991) Implications of DNA replication for eukaryotic gene expression. *J Cell Sci* **99**: 201-206

Wong HC, Fear AL, Calhoon RD, Eichinger GH, Mayer R, Amikam D, Benziman M, Gelfand DH, Meade JH, Emerick AW, Bruner R, Ben-Bassat A, Tal R (1990) Genetic organisation of the cellulose synthase operon in *Acetobacter xylinum*. *Proc Natl Acad Sci USA* **87**: 8130-8134

Woudt LP, Pastink A, Kempers-Veenstra AE, Jansen AEM Mager WH, Planta RJ (1983) The genes coding for histone H3 and H4 in *Neurospora crassa* are unique and contain intervening sequences. *Nucl Acids Res* **11**: 5347-5360

Wright HT, Sandrasegaram G, Wright CS (1991) Evolution of a family of N-acetylglucosamine binding proteins containing the disulphide-rich domain of wheat germ agglutinin. *J Mol Evol* **33**: 283-294

Wu L, Ueda T, Messing J (1993) 3'-end processing of the maize 27 kDa zein mRNA. *Plant J* **4**: 535-544

Wyatt RE, Nagao RT, Key JL (1992) Patterns of soybean proline-rich protein gene expression. *Plant Cell* **4**: 99-110

Xu W, Purugganan MM, Polisensky DH, Antosiewicz DM, Fry SC, Braam J (1995) *Arabidopsis TCH4*, regulated by hormones and the environment, encodes a xyloglucan endotransglycosylase. *Plant Cell* **7**: 1555-1567

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:103-119.

Yatsu LY (1983) Morphological and physical effectas of colchicine treatment on cotton (*Gossypium hirsutum* L.) fibres. *Text Res J* **53**:515-519

Yatsu LY, Jacks TJ (1981) An ultrastructural study of the relationship between microtubules and microfibrils in cotton (*Gossypium hirsutum* L.) cell wall reversals. *Amer J Bot* **68**: 771-777

Ye Z-H, Varner JE (1991) Tissue-specific expression of cell wall proteins in developing soybean tissues. *Plant Cell* **3**: 23-37

Ye Z-H, Song Y-R, Marcus A, Varner JE (1991) Comparitive localisation of three classes of cell wall proteins. *Plant J* **1**: 175-183

Young RA (1986) Structure, swelling and bonding of cellulose fibres. *In* RA Young, RM Rowell (eds) *Cellulose, Structure, Modification and Hydrolysis*. John Wiley and Sons, NY, pp 91-128

Zhou J, Goldsbrough PB (1995) Structure, organisation and expression of the metallothionein gene family in *Arabidopsis*. *Mol Gen Genet* **248**: 318-328

Zhu J-K, Damsz B, Kononowicz AK, Bressan RA, Hasegawa PM (1994) A higher plant extracellular vitronectin-like adhesion protein is related to the translational elongation factor-1α. *Plant Cell* **6**: 393-404