

THE ISOLATION AND CHARACTERISATION OF FATTY ACID DESATURASE GENES IN COTTON

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
at the University of Sydney

by

Qing Liu (M.Ag.Sci. Adelaide)

March 1998

ABSTRACT

Cottonseed oil is an important commodity both in Australia and worldwide. Cottonseed oil contains about 22-26% saturated fatty acid, mainly palmitic acid; 15-20% monounsaturated fatty acid (oleic acid), 50-58% polyunsaturated fatty acid, mainly linoleic acid. The $\Delta 9$ stearyl-ACP desaturase and microsomal ω -6 desaturase are the two major enzymes responsible for the synthesis of unsaturated fatty acid in cottonseed oil. In order to facilitate the genetic engineering of cottonseed oil towards the improved oil stability and nutritional values for human consumption, it is necessary to isolate and characterise the key genes encoding the major fatty acid desaturase enzymes in cottonseeds.

A full length cDNA encoding the $\Delta 9$ stearyl-ACP desaturase, pgh9-1, was isolated from a λ cDNA library of immature embryos from *G. hirsutum*. Similar to other plant $\Delta 9$ stearyl-ACP desaturase cDNAs, pgh9-1 encodes a polypeptide of 397 amino acids with the first 32 amino acid residues being a transit peptide. Northern blot analysis using the pgh9-1-specific probe indicated that the expression of this gene was highly induced during cottonseed embryo development, but no expression was observed in leaf tissues. Therefore, pgh9-1 is presumed to be responsible for the introduction of first double bond into stearyl-ACP during fatty acid synthesis and oil accumulation in cottonseeds.

The structural gene, Gh9-1G, corresponding to pgh9-1 was isolated from a *G. hirsutum* genomic DNA library. It consists of three exons and two introns. In addition to the TATA box and CAAT box (Messing et al., 1983), several previously identified seed-specific promoter motifs, including ABREs-like (Guiltinan et al., 1990), RY-like (Bäumlein et al., 1986) and a AGCCCG motif (Chen et al., 1988), were also found in the 5' flanking region of the Gh9-1G. Interestingly, a region homologous to the cotton

2S storage protein Mat5-A promoter was located near the ABREs-like and RY-like motifs in Gh9-1G. It is speculated that this region may contain regulatory elements which can enhance or refine the seed specificity. Southern blot analysis revealed that the Gh9-1G is a member of a multi-gene family comprised of at least 6-8 genes (or copies) per diploid genome in *Gossypium*. The diploid *Gossypium* species contains at least one gene (copy), and the allotetraploid contains at least two copies, each of which represents its subgenomic origin.

Two cDNAs encoding *G. hirsutum* microsomal ω -6 desaturase were also isolated from a λ cottonseed cDNA library, and named as pghD12-1 and pghD12-2, respectively. The sequence comparison between pghD12-1 and pghD12-2 revealed that they share 70.7% sequence identity in the coding region, but with unique 5' and 3' UTRs. This may indicate that they have evolved independently since their divergence from a common ancestral gene.

Northern blot analysis revealed that the pghD12-1 gene was not expressed in leaf tissue, but specifically activated in developing cotton embryos. It is likely, therefore, that pghD12-1 is responsible for the introduction of the second double bond into the fatty acid chain in cottonseed oil. In contrast, pghD12-2, has very low expression in both cotton leaf tissues and developing embryos. It may not play a major role in the desaturation of cottonseed storage lipids and membrane lipids. Southern blot analyses indicate that in tetraploid cotton, there are at least five closely related genes (copies) among which at least three genes (copies) were from its putative A-genome progenitor and two from D-genome progenitor. Southern blot analysis also indicated that the pghD12-1 and pghD12-2 are nonallelic and there are at least two copies of each gene in the tetraploid cotton species and at least one copy of each in diploid species.

The genomic structure of pghD12-1 has also been studied and it was found that pghD12-1 gene contains a large intron (1.1 kb) in its 5' untranslated region (UTR). This intron has been sequenced from a number of *Gossypium* species, including both of the A-genome species, three representatives of the D-genome group, two AD-allotetraploids

and all seventeen Australian native *Gossypium* species. Phylogenetic analysis of the intron sequences using distance matrix produced topologies largely consistent with previous phylogenetic studies of *Gossypium*. However, it also revealed some interesting observations in the evolutionary history of pghD12-1 gene. Firstly, in the initial polymerase chain reaction (PCR) amplification of the pghD12-1 intron from allotetraploids, only the sequences from A-subgenome were amplified. Evidences were shown that the primer mismatch in the 5' UTR of D-subgenome pghD12-1 gene contributed to the PCR failure. Southern blot analysis also indicated that the 3' UTR in the D-subgenome pghD12-1 gene has undergone relatively rapid sequence alterations since the interspecific hybridisation and polyploidisation occurred 1-2 millions years ago. However, partial sequence of the D-subgenome pghD12-1 gene isolated from allotetraploid *Gossypium* showed high homology with the corresponding region in D-genome diploid. It is, therefore, speculated that the rapid sequence alteration occurred mainly in the UTR regions. Secondly, incongruent with previous phylogenetic studies, a G-genome species, *G. bickii* was placed in the C-genome clade. It has been previously demonstrated that *G. bickii* has biphyletic history and contains a G nuclear genome and a C chloroplast genome (Wendel and Albert, 1992). The phylogenetic information inferred by the DNA sequence of pghD12-1 intron probably represents the first evidence that *G. bickii*'s capture of the C chloroplast genome was also accompanied by nuclear introgression. Nuclear introgression was also detected in one accession of *G. australe* from its sister taxa, *G. nelsonii*.

DECLARATION

I HEREBY DECLARE that the work presented in this thesis has been carried out by myself and does not incorporate any material previously submitted for another degree in any university. To the best of my knowledge and belief, it does not contain any material previously written or published by another person, except where due reference is made in the text. I am willing to make the thesis available for photocopy and loan if it is accepted for the award of the degree.


Qing Liu

ACKNOWLEDGMENTS

I am indebted to my supervisors, Drs S.P. Singh, P.J. Sharp, A.G. Green and Prof. D.R. Marshall for their continuous support, supervision and inspiration during the course of this study.

I wish to thank Dr Greg Constable in Australian Cotton Research Institute, Narrabri, NSW for providing cottonseeds for oil and fatty acid analysis. I also wish to thank Prof. Derrick Balnov in Sydney University for allowing me access to the GLC equipment in his laboratory.

Other people have helped me in innumerable ways, and of these I particularly wish to acknowledge (in alphabetical order): Mick Ayliffe, Curt Brubaker, Peter Flint, Zhongyi Li, Danny Llewellyn, Bruce Lyon, Odile Moullet, Tony Pryor, Sadequr Raman, Peter Ryan and Peter Stoutjesdijk.

Particular thanks to the members of the Pryor/Singh laboratories in CSIRO Plant Industry for constructive discussions, social interaction and general tolerance -especially Sue McKinney for her constant assistance, advice and friendship.

I gratefully acknowledge the receipt of the Postgraduate Scholarship from Australian Cotton Research and Development Corporation. I am also grateful to all the support from the Program P in CSIRO Plant Industry.

I would like to express my deepest appreciation to my parents for their patience and support throughout my study and all these years in Australia. At last, not the least, I wish to thank my wife Xiaoying for her understanding, support and love.

LIST OF CONTENTS

Title Page	i
Abstract	ii
Declaration.....	v
Acknowledgments.....	vi
List of Contents.....	vii
List of Tables.....	xiii
List of Figures	xiv

CHAPTER ONE

LITERATURE REVIEW

1.1	Increasing Demands for Vegetable Oil	1
1.2	Cottonseed Oil as an Important Commodity both in Australia and Worldwide.....	1
1.3	Targets for Genetic Modification of Cottonseed Oil for Food Applications	3
1.3.1	High Oleic Cottonseed Oil	3
1.3.2	High Stearic Cottonseed Oil and Cocoa Butter Substitute	3
1.4	Strategies for Modifying Plant Lipid Composition	4
1.5	Biochemistry of Seed Lipid Biosynthesis.....	5
1.5.1	Fatty Acids in Major Oil Crops.....	6
1.5.2	Plant <i>de novo</i> Fatty Acid Synthesis.....	7
1.5.3	Termination of the Fatty Acid Chain and Two Parallel Pathways for Lipid Biosynthesis	10
1.5.4	Extra-plastidial Metabolism of Fatty Acids	10
1.5.5	Cloning of Plant FAS and Acyltransferase Genes	13
1.5.6	Desaturation of Fatty Acids.....	16
1.5.6.1	$\Delta 9$ Stearoyl-ACP Desaturase and Other Acyl-ACP Desaturases	16
1.5.6.1.1	$\Delta 9$ Stearoyl-ACP Desaturase.....	17
1.5.6.1.2	Other Acyl-ACP Desaturases.....	19
1.5.6.2	Membrane-bound Desaturases	20
1.5.6.2.1	Cloning of ω -6 Desaturase Genes	21
1.5.6.2.2	Cloning of ω -3 Desaturase Genes	22
1.5.6.2.3	Cloning of Other Membrane-bound Desaturase Genes.....	23

	1.5.6.2.4 Conserved Domains in Membrane-bound Desaturases.....	24
1.6	Progress in Genetic Engineering of Fatty Acid Composition in Oilseed Crops	24
1.6.1	Introduction of Novel Genes	25
1.6.2	Downregulation of Unwanted Traits	25
1.7	Synopsis of the Work Embodied in This Thesis.....	27

CHAPTER TWO

GENERAL MATERIALS AND METHODS

2.1	Growth of Plant Materials	30
2.2	Growth of Bacteria	30
2.3	Transformation of <i>E. coli</i> with Plasmids by Electroporation	30
2.4	DNA and RNA Isolation and Purification	31
2.4.1	Plasmid DNA Miniprep	31
2.4.2	Purification of Plasmid DNA for DNA Sequence Determination	31
2.4.3	Isolation of Genomic DNA from Plant Materials	32
2.4.4	Purification of Genomic DNA by CsCl Gradient	33
2.4.5	Isolation of RNA	34
2.5	DNA and RNA Gel Electrophoresis and Blotting	35
2.5.1	Preparation of a DNA-Minigel	35
2.5.2	DNA Southern Transfer	35
2.5.3	RNA Northern Transfer	36
2.6	Preparation of ³² P-Labelled DNA Probes	36
2.7	Hybridisation and Autoradiography	37
2.8	Construction of a Cottonseed cDNA Library	37
2.8.1	mRNA Purification	37
2.8.2	cDNA Synthesis	38
2.8.3	Construction of a cDNA Library	39
2.8.3.1	Ligation and Packaging	40
2.8.3.2	Titering of cDNA Library	40
2.8.4	Amplification of the Cottonseed cDNA Library	41
2.9	Screening of a cDNA Library	42
2.9.1	Plating of a cDNA Library	42
2.9.2	Screening by DNA Hybridisation	42
2.9.3	<i>In vivo</i> Excision of Phagemid	43
2.10	Making of Gene Constructs	44

CHAPTER THREE

OIL CONTENT AND FATTY ACID COMPOSITION IN DEVELOPING AND MATURE COTTONSEEDS

3.1	Introduction	47
3.2	Materials and Methods	48
3.2.1	Plant Materials	48
3.2.2	Lipid Extraction	49
3.2.3	Fatty Acid Analysis by Gas-Liquid Chromatography (GLC)	50
3.2.4	Statistical Analysis	50
3.3.	Results and Discussion	51
3.3.1	Oil Content in the Major Australian Cotton Cultivars and Some Other Genotypes and Species	51
3.3.2	The Changes of Fatty Acid Composition in Developing Cottonseeds	57
3.4	Summary.....	59

CHAPTER FOUR

ISOLATION AND CHARACTERISATION OF Δ^9 STEAROYL-ACP DESATURASE CDNA AND ITS GENOMIC STRUCTURAL GENE FROM *G. hirsutum*

4.1	Introduction.....	60
4.2	Materials and Methods	61
4.2.1	Oligonucleotide Design and Reverse Transcriptase-PCR (RT-PCR).....	61
4.2.2	Isolation of a cDNA from a Cottonseed cDNA Library	62
4.2.3	Isolation of the Structural Gene from a Cotton Genomic DNA Library	63
4.2.4	RNA Northern Blot Analysis	64
4.2.5	Genomic DNA Southern Blot Analysis	65
4.3	Results	65
4.3.1	Cloning and Sequence Analysis of a cDNA pgh9-1	65
4.3.2	Gene Expression of pgh9-1 Revealed by Northern Blot Analysis	70
4.3.3	The Primary Structure of A Δ^9 Stearoyl-ACP Desaturase Gene from <i>G. hirsutum</i>	72
4.3.4	The Δ^9 Stearoyl-ACP Desaturase Gene Family.....	78
4.4	Discussion	82

CHAPTER FIVE

ISOLATION AND CHARACTERISATION OF cDNA CLONES ENCODING MICROSOMAL ω -6 FATTY ACID DESATURASE IN *G. hirsutum*

5.1	Introduction	90
5.2	Materials and Methods	92
5.2.1	Isolation of cDNA Clones from a Cottonseed λ cDNA Library	92
5.2.2	Amplification of cDNA ends from λ cDNA Library DNA	92
5.2.3	Primer-adaptor Mediated PCR	93
5.2.4	RNA Northern Blot Analysis	93
5.2.5	Genomic DNA Southern Blot Analysis	94
5.3	Results	94
5.3.1	Isolation and Sequence Analysis of cDNAs Encoding Microsomal ω -6 Desaturase from <i>G. hirsutum</i>	94
5.3.1.1	Isolation of Two 5' Truncated cDNAs from cDNA Library	94
5.3.1.2	PCR Amplification of 5' end of pghD12-1 from cDNA Library	95
5.3.1.3	Amplification of 5' Upstream Sequence of pghD12-2 from Genomic DNA	97
5.3.2	Evaluation of Gene Expression by RNA Northern Blot Analysis	99
5.3.3	Genomic Southern Blot Analysis in <i>Gossypium</i> Species	101
5.4	Discussion	104
5.4.1	Sequence Features of cDNAs Encoding Microsomal ω -6 Desaturases in <i>G. hirsutum</i>	104
5.4.2	Comparison of the Deduced Polypeptide Sequences of <i>G. hirsutum</i> ω -6 Desaturase and Other Plant Membrane-bound Fatty Acid Desaturases	106
5.4.3	Hydropathy Analysis of pghD12-1 and pghD12-2	110
5.4.4	PCR Strategies for Rapid Amplification of 5' cDNA Ends	113
5.4.5	Gene Expression Patterns of Cotton Microsomal ω -6 Desaturase	114
5.4.6	Estimation of Gene Copy Number and the Size of Microsomal ω -6 Desaturase Multigene Family	115

CHAPTER SIX

EVOLUTIONARY STUDIES OF PGHD12-1 GENE IN *Gossypium* SPECIES ON THE BASIS OF ITS 5'UTR INTRON DNA SEQUENCE

6.1	Introduction	117
6.2	Materials and Methods	119
6.2.1	Plant Materials and DNA Extractions.....	119
6.2.2	Adapter-Primer Mediated PCR	121
6.2.3	PCR Amplification and Cloning of 5' UTR Introns from <i>Gossypium</i> Species	121
6.2.4	Sequence Alignment and Analysis	122
6.2.5	Amplification of the Corresponding Intron from <i>A. thaliana</i> and <i>G. max</i>	123
6.2.6	High Stringency Southern Blot Analysis	123
6.2.7	Construction of a 35S-intron-GUS Cassette	124
6.3	Results	125
6.3.1	The Localisation of pghD12-1 Intron from <i>G. hirsutum</i>	125
6.3.2	The Amplification of the pghD12-1 Intron from Other <i>Gossypium</i> Species.....	127
6.3.3	DNA Sequences of the Corresponding Intron from <i>A. thaliana</i> and <i>G. max</i>	142
6.3.4	PCR Amplification of D-subgenome Intron from Tetraploid <i>Gossypium</i> Species	143
6.3.5	Southern Blot Analysis at High-Stringency Conditions	151
6.3.6	Construction of an 35S-Intron-GUS Fusion System	153
6.4	Discussion	154
6.4.1	Sequence Characteristics of the Microsomal ω -6 Fatty Acid Desaturase Intron	154
6.4.2	The Sequence Variations in the N-terminal of pghD12-1 ORF Among <i>Gossypium</i> Species	156
6.4.3	The Sequence Composition of a Microsatellite Within the pghD12-1 Intron	158
6.4.4	Alterations of DNA Sequence in the D-subgenome pghD12-1 Gene in Tetraploid Cottons	160
6.4.5	Nuclear Introgression in Australian <i>Gossypium</i> Species	163
6.4.6	The Phylogenetic Position of <i>G. gossypioides</i> pghD12-1 Intron	166
6.4.7	The Potential Gene Regulation by the pghD12-1 Intron	167
6.5	Conclusions	106

CHAPTER SEVEN

GENERAL DISCUSSION

7.1	Genetic Engineering of Cottonseed Oil.....	170
7.1.1	High Stearic Cottonseed Oil	171
7.1.2	Low Palmitic Cottonseed Oil	172
7.1.3	High Oleic Cottonseed Oil	172
7.1.4	Cocoa Butter Substitute	173
7.1.5	A Single Dominant Fatty Acid in Cottonseed Oil	173
7.2	Modification of Fatty Acid Desaturase by Protein Engineering	174
7.2.1	Modification of Substrate and Positional Specificities of Δ^9 Stearoyl-ACP Desaturase	174
7.2.2	Potential New Functions of Microsomal ω -6 Desaturase	175
7.3	Optimisation of Transgene Regulation	176
7.3.1	Use of a Seed Storage Protein Promoter	176
7.3.2	Isolation and Potential Utilisation of a Fatty Acid Desaturase Gene Promoter	176
7.3.3	Identification of Transcription Factor and Environmental Influences	178
7.3.4	Importance of UTRs in the Transgene Expression	179
7.3.5	Potential Enhancing Effect of the pghD12-1 Intron	179
7.4	Intron Gain or Intron Loss from the Microsomal ω -6 Desaturase Genes in Cotton	180
7.5	Evolution of Fatty Acid Desaturase Genes in Allotetraploid Cotton	181
7.6	Potential Utilisation of Intron Sequences in the Phylogenetic Analysis of <i>Gossypium</i> Species	183
	REFERENCES	184

LIST OF TABLES

Table 1.1	Australian Oilseed Production 1993/94	2
Table 1.2	Structure of the Major Fatty Acids in Major Oilseeds	6
Table 1.3	A Summary of Gene Cloning of Plant Protein and Key FAS Enzymes Involved in the Fatty Acid Synthesis and Acyl Transfer	13
Table 3.1	Lipid Content and Fatty Acid Analysis in Cottonseed Oil	55
Table 3.2	Correlation Coefficients of Fatty Acid Composition of Cottonseed Oil	56
Table 4.1	Sequence Similarities between pgh9-1 and Other cDNAs for $\Delta 9$ Stearoyl-ACP Desaturase	69
Table 4.2	The AT Content of intron vs. ORF in $\Delta 9$ Stearoyl-ACP Desaturase Gene in Three Plant Species (<i>B. napus</i> , <i>S. indicum</i> and <i>G. hirsutum</i>).....	76
Table 6.1	A List of <i>Gossypium</i> Accessions from Which the pghD12-1 Intron Was Isolated and Sequenced	119
Table 6.2	A Summary of the Features of the Microsomal ω -6 Desaturase Intron from <i>G. hirsutum</i> , <i>G. max</i> and <i>A. thaliana</i>	154
Table 6.3	A Summary of the Microsatellite (GAA) _n Pattern in the pghD12-1 Intron from <i>Gossypium</i> Species	159

LIST OF FIGURES

Fig. 1.1	Schematic Overview of Fatty Acid Synthesis in Plants	9
Fig. 1.2	A Simplified Diagram of the Scheme of Phospholipid and Triacylglycerol Synthesis in Plants	12
Fig. 2.1	Plasmids Used to Make the Sense and Antisense Gene Constructs of Cotton Fatty Acid Desaturase Gene Driven by a Lectin Promoter	46
Fig. 3.1	A Typical Chromatography Chart of the Fatty Acid Analysis of Cottonseed Oil in the Current Study	53
Fig. 3.2	Seed Oil Accumulation and Changes of Fatty Acid Composition in Developing Cottonseed Embryos	58
Fig. 4.1	Nucleotide and Deduced Amino Acid Sequences of a cDNA, Designated as pgh9-1, Encoding for the $\Delta 9$ Stearoyl-ACP Desaturase of <i>G. hirsutum</i>	66
Fig. 4.2	Northern Blot Analysis of Developing Embryos and Leaf Tissues from <i>G. hirsutum</i> cv Deltapine-16	71
Fig. 4.3	Gene Structure and Nucleotide Sequence of a $\Delta 9$ Stearoyl-ACP Desaturase Gene, Gh9-1G, from <i>G. hirsutum</i>	74
Fig. 4.4	Gap Analyses of Two Homologous Regions (A and B) in the Promoter of Gh9-1G and Those of Cotton 2S Storage Protein (Mat5-A) Gene	78
Fig. 4.5	Southern Blot Analysis of $\Delta 9$ Stearoyl-ACP Desaturase Genes in <i>Gossypium</i> Species	80
Fig. 4.6	A Dendrogram Showing the Relationship between Various Acyl-ACP Desaturases According to Their Nucleotide Sequence Similarities in the ORF	84
Fig. 5.1	Nucleotide and Deduced Amino Acid Sequences of pghD12-1	96
Fig. 5.2	Nucleotide and Deduced Amino Acid Sequences of pghD12-2	98
Fig. 5.3	Northern Blot Analysis of Developing Embryos and Leaf Tissues from <i>G. hirsutum</i> cv Deltapine-16	100
Fig. 5.4	Southern Blots of Cotton Genomic DNAs Probed with Cotton Microsomal ω -6 Desaturase cDNA Fragments	102
Fig. 5.5	Alignment of Nucleotides Surrounding the Translation Initiation Codon ATG of the Two Full Length cDNAs Encoding for Cotton Microsomal ω -6 Desaturases and the Plant Consensus Motif for Translation Initiation	104
Fig. 5.6	Comparison of the Deduced Amino Acid Sequences of Five Microsomal ω -6 Desaturases and Other Membrane-bound Fatty Acid Desaturases Including Microsomal ω -3 Desaturase and	

	Plastidial ω -6 Desaturase	109
Fig. 5.7	Hydropathy Analysis of the Microsomal ω -6 Desaturase in <i>G. hirsutum</i> Using Kyte-Doolittle Method	112
Fig. 6.1	Comparison of Nucleotide Sequences between the cDNA pghD12-1 and the PCR Fragment-A Amplified from <i>G. hirsutum</i> Genomic DNA with Primer <i>d12A4</i> and the EcoRI Primer	125
Fig. 6.2	Genomic Structure of the Microsomal ω -6 Desaturase Gene Corresponding to cDNA pghD12-1 in <i>G. hirsutum</i>	126
Fig. 6.3	Sequence Alignment of the 5' UTR Intron and a Portion of Its 3' Flanking Exon Region of pghD12-1 Gene in 32 Accessions of 24 <i>Gossypium</i> Species.....	129
Fig. 6.4	A Phylogenetic Tree of 24 <i>Gossypium</i> Species Based on the Entire 5' UTR Intron and a Portion of 3' Flanking Exon of Microsomal ω -6 Desaturase pghD12-1	141
Fig. 6.5	Nucleotide Sequences of the 5' UTR Intron of FAD2 Gene in <i>A. thaliana</i> (A) and Intron 1 of FAD2-1 in <i>G. max.</i> (B)	142
Fig. 6.6	PCR Amplification of the pghD12-1 Intron of D-subgenome Origin from Allotetraploids	144
Fig. 6.7	The Subgenomic Intron Sequences of pghD12-1 Gene from Allotetraploids and Those Closely Related to the Putative Diploid Progenitors.....	147
	(A). DNA Sequence Alignment.	147
	(B). A Distances Based Phylogenetic Tree.	150
Fig. 6.8	Southern Blot Analysis of EcoRI- and HindIII-digested Genomic DNAs from <i>G. barbadense</i> , <i>G. hirsutum</i> , <i>G. herbaceum</i> , <i>G. raimondii</i> and <i>G. robinsonii</i>	152
Fig. 6.9	A 35S-intron-GUS chimeric Construct	153
Fig. 6.10	An Alignment of the Deduced Amino Acid Sequences in the N-terminal Portion (48 Amino Acid Residues) of the pghD12-1 Gene from 32 Accessions of 24 <i>Gossypium</i> Species	157

Chapter One

Literature Review

1.1 INCREASING DEMANDS FOR VEGETABLE OIL

Plant storage lipids or oils are of major economic importance, representing a source for practically all plant-derived fats in our diet and the basis for the oleo-chemical industry. World demand for vegetable oils continues to grow as a result of higher levels of fat consumption in developing countries, caused by higher levels of disposable income and due to the healthier image of vegetable oils as compared to animal fat in developed countries. Since 1992/93, the world production of major oilseeds has grown by over 13% to some 256 million metric tonnes (mt) in 1996 (Hedley, 1997). The human consumption of major vegetable and marine oils increased by some 19% between 1992/93 and 1995/96 and was forecast to increase by a further 2% in 1996/97 (Hedley, 1997).

1.2 COTTONSEED OIL AS AN IMPORTANT COMMODITY BOTH IN AUSTRALIA AND WORLDWIDE

Among the oil seeds, cottonseed is considered to be an important commodity because of its use in food and animal feed throughout the world. The cottonseed contains approximately 23-26% oil and 26-28% protein on a moisture, lint-free seed basis (Cherry, 1984). The world production of cottonseed oil in 1994/95 was 3.8 million mt

ranking sixth behind soybean, oil palm, rapeseed, sunflower and groundnut (Oil World Annual, 1996). Global cottonseed production will expand in the future as a result of increased acreage and higher yields (Baize, 1997).

In Australia, the production of all oilseeds is approximately 1 million mt per annum, about half of which is cottonseed (Table 1.1). While Australia imports soybean, sunflower, olive, palm and coconut oils, it is a surplus producer of cottonseed.

Table 1.1 Australian Oilseed Production 1993/94 (Burden, 1995)

OILSEED	PRODUCTION ('000 mt)
Canola	286
Cottonseed	466
Sunflower	123
Soybeans	79
Safflower	18
Peanuts (non-edible)	5
Maize germ	3
High oleic sunflower	4
Linola	7
TOTAL	991

Cottonseed oil, however, is a by-product of the more valuable fibre crop, and its availability is mainly influenced by the demand for cotton fibre. During 1981-90 the cotton fibre contributed 80.7% of the economic value of the cotton crop, while the oil and meal component accounted for only 8.1% and 7.9% of the value, respectively (Stanton and Blumenfeld, 1992).

Currently cottonseed oil is mainly used for human consumption. However, it has to be blended or chemically modified before use because its fatty acid composition is non-optimal. There is, therefore, increasing demand for upgrading cottonseed oil *in planta*.

The ideal profile of the modified cottonseed oil should be high oleic, high stearic and possibly be used as a cocoa butter substitute.

1.3 TARGETS FOR GENETIC MODIFICATION OF COTTONSEED OIL FOR FOOD APPLICATIONS

1.3.1 High Oleic Cottonseed Oil

Most of the vegetable oils, including cottonseed oil, produced for commercial food applications are hydrogenated to increase their oxidative stability during storage or frying, or to make the oil solid for edible spreads and margarines. An additional side reaction that occurs during hydrogenation results in the formation of the *trans* isomer of oleic acid, i.e. eliadic acid, which is not normally found in natural plant oils. The consumption of the *trans* isomer in the diet has been regarded as a possible risk factor for coronary heart disease (Khosla and Hayes, 1996). Another disadvantage of hydrogenation is the additional cost added to the price of the oil. Thus, it is desirable to find an alternative to vegetable oil hydrogenation, such as the production of plant oils which are naturally high in oleic acid. This goal could be accomplished in cottonseed *via* the reduction of the microsomal ω -6 desaturase levels using antisense suppression or cosuppression, as has been achieved in rapeseed and soybean (Hitz et al., 1995; Kinney, 1996).

1.3.2 High Stearic Cottonseed Oil and Cocoa Butter Substitute

As it will be discussed in the following sections, stearate does not accumulate in cottonseed and the vast majority of oilseeds. However, the production of high stearic plant oil is highly desirable because of its potential to be used in margarine production, confectionary foods (as a cocoa butter substitute) and in the pharmaceutical and cosmetic industries (Töpfer et al., 1995). Cocoa butter is the most highly valued plant lipid and commands a large price premium relative to other vegetable fats and oils. Currently, the seeds of the cocoa tree, *Theobroma cacao* L., are the sole source of cocoa butter, hence

the cocoa butter is often in short supply and the price can fluctuate considerably, being strongly influenced by the cocoa cultivation and harvest in the limited cocoa growing areas of the world.

In cocoa butter, the major species of triacylglycerol is mainly composed of saturated fatty acid at positions sn-1 (palmitate) and sn-3 (stearate), and with oleate at position sn-2 (Fincke, 1976; Gunstone et al., 1986). These acyl features of the so-called POS type triacylglycerol give cocoa butter its unique physical properties of sharp melting point and mouth feeling and therefore being used to make high-value products such as chocolate. A challenging target for the genetic modification of edible oils is the production of a cheaper alternative to cocoa butter for use in chocolate and confectionary industries. The major difference between the cocoa butter and the relatively inexpensive vegetable oils, such as cottonseed oil, is the degree of unsaturation and the arrangement of the fatty acids. It has already been shown that given the right proportions of palmitate, stearate and oleate both safflower and turnip-rape cell-free microsomes can construct cocoa butter-like fats *in vitro* (Bafor et al., 1990). This clearly indicates the likelihood that POS triacylglycerols would be synthesised in these plants if the proportion of fatty acids were appropriate. Thus, engineered vegetable oil containing predominantly palmitate, oleate and stearate in equal proportion may give rise to POS triacylglycerol.

1.4 STRATEGIES FOR MODIFYING PLANT LIPID COMPOSITION

Modification of the fatty acid composition may be accomplished in different ways (for review, see Röbbelen et al., 1989). Natural variability for these traits has been observed in safflower (Knowles, 1989) and rapeseed (Stefansson et al., 1961). Additional variability has also been obtained by mutagenesis or genetic engineering (Ohlrogge et al., 1991). The induction of genetic variability by mutagenesis has made possible the identification and isolation of mutants with altered fatty acid composition in the seed lipids of commercial oilseed crops (Ratray, 1991). The development of novel linseed oil with significantly reduced α -linolenic acid (Green and Marshall, 1984; Green, 1986a,b),

development of high stearate soybean (Graef et al., 1985), high oleate sunflower (Miller et al., 1987), high oleate safflower (Purdy, 1986) and very low erucic acid (<1%) canola varieties (Harvey and Downey, 1964) represent successful examples of conventional genetic approaches *via* mutagenesis or selection.

However, conventional breeding is generally very time-consuming and labor-intensive and involves elaborate screening procedures in order to detect the desired biochemical traits. It has also been limited by the lack of useful genetic variations within a species and the non-specificity of mutagenesis. Recent progress in understanding the genetic and biochemical factors regulating plant lipid composition raises the possibility that this limitation may be alleviated by using cloned genes to create genetically modified plants that produce 'designer' oils, in which the fatty acid composition is precisely tailored to suit particular uses. The first food product made from transgenic oilseed crop is now commercially available in the USA, i.e. high laurate canola oil (Voelker, 1995). In addition, transgenic lines of high oleate soybeans and high stearic canola have also been developed (Kinney, 1995; Knutzon et al., 1992b).

1.5 BIOCHEMISTRY OF SEED LIPID BIOSYNTHESIS

A prerequisite to any genetic modification of seed oil composition is a detailed knowledge of the metabolic pathways involved in the fatty acid biosynthesis. This is a complex, highly regulated and multicompartmented series of biochemical reactions. A thorough biochemistry knowledge will provide the basis for identification, and subsequent isolation and purification of enzymes, thereby enabling the cloning of appropriate genes for genetic engineering. This chapter will place emphasis on the biochemistry and molecular cloning of fatty acid desaturase genes following a general description of *de novo* fatty acid synthesis and their assembly into triacylglycerol.

1.5.1 Fatty Acids in Major Oil Crops

Plant seed oils usually contain over 95% neutral storage lipids in the form of triacylglycerol (TAG), i.e., glycerol esters of saturated or unsaturated fatty acids. The properties of the oil, and hence its use and value, are determined by the fatty acid composition and, to some extent, the positional specificity of the fatty acids attached to the glycerol backbone (Gunstone and Norries, 1983). There are over 300 different fatty acids in nature, including fatty acids with different chain length, such as the 10:0 fatty acid in elm seed oil to erucic acid (22:1) in rapeseed oil; and fatty acids with different double bond position or modifying groups, such as hydroxylated fatty acid in castor bean to cyclopropene fatty acids in *Malvaceae* seeds, etc. In contrast to the extreme variability of fatty acid composition in the wide range of plant seed oil found across seed plants as a whole, the six major oil crops (soybean, oil palm, rapeseed, sunflower, cottonseed, and groundnut) which account for 84% of worldwide vegetable oil production (Lühs and Friedt, 1994), consists mainly of palmitic, stearic, oleic, linoleic and α -linolenic acids (Table 1.2).

Table 1.2 Structure of the Major Fatty Acids in Major Oilseeds

COMMON NAME	SYMBOL	STRUCTURE
Palmitic acid	16:0	$\text{CH}_3(\text{CH}_2)_{14}\text{COOH}$
Oleic acid	18:1 ^{Δ9}	$\text{CH}_3(\text{CH}_2)_7\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$
Linoleic acid	18:2 ^{Δ9,12}	$\text{CH}_3(\text{CH}_2)_4\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$
α -Linolenic acid	18:3 ^{Δ9,12,15}	$\text{CH}_3\text{CH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$

1.5.2 Plant *de novo* Fatty Acid Synthesis

The broad aspects of fatty acid biosynthesis, including the sequential desaturation was first suggested by Harris and James (1965). Over the last decade, knowledge of the individual enzymatic steps involved in *de novo* fatty acid synthesis has dramatically increased due to the purification of proteins involved and more recently the isolation of their encoding genes (Töpfer et al., 1995).

The *de novo* synthesis of saturated fatty acids and their initial desaturation are carried out exclusively in the stroma of plastids while further modifications of fatty acyl molecules and TAG assembly are localised in the cytoplasm (Fig. 1.1 & Fig. 1.2). As illustrated in Fig. 1.1, the pathway of fatty acid biosynthesis involves the repeated addition of two-carbon units derived from acetyl-CoA to initiate and then extend an acyl group to 16 or 18 carbons in length. The main enzymes involved in this synthesis are acetyl-CoA carboxylase (ACCase) and fatty acid synthase (FAS). ACCase is a biotinylated protein catalysing the first committed step of fatty acid biosynthesis producing malonyl-CoA from acetyl-CoA. The malonyl group is then transferred from CoA to a protein cofactor, acyl carrier protein (ACP). From this point on, all the reactions in the rest of the pathway occur on an ACP substrate until the 16 or 18 carbon product is ready for transfer to glycerolipids or export from the plastid. The chain elongation occurs by the sequential condensation of two-carbon units, derived from malonyl CoA, to the acyl chain catalysed by FAS. There are two types of FAS, type I and II in nature. While type I (FASI) is found in animals, yeast and mycobacteria, type II (FASII) is found in *E. coli*, cyanobacteria and plants. FASII is a collection of individual enzyme activities catalysing the conversion of acetyl-CoA and malonyl-CoA to 16:0 and 18:0. The condensation reactions of FASII enzymes are performed by three different β -ketoacyl-ACP synthases (KAS), i.e. KASI, II and III, which are distinguished by their substrate specificity. The first condensation reaction is catalysed by KASIII using acetyl-CoA and malonyl-ACP (Jaworski et al., 1989; Walsh et al., 1990; Schuch et al., 1993). KASI is then used to extend the acyl chain from C4 up to C16 (Shimakata and Stumpf, 1982). In the final

condensation reaction mediated by KASII, C16-ACP is elongated to C18-ACP (Shimakata and Stumpf, 1982). The initial product of each condensation reaction is 3-ketoacyl-ACP. Three additional reactions occur after each condensation to form a saturated fatty acid. The 3-ketoacyl-ACP is first reduced at the carbonyl group by the enzyme 3-ketoacyl-ACP reductase and then dehydrated by hydroxyacyl-ACP dehydrase. The cycle of fatty acid synthesis is then completed by the enzyme enoyl-ACP reductase, which uses NADH and NADPH to reduce the *trans*-2 double bond to form a saturated fatty acid. In most lipid-accumulating tissues in plants, the end products of the FAS reaction are usually stearyl-ACP and to a lesser extent, palmitoyl-ACP. Oleate (C18:1), the major monounsaturated fatty acid of plant oils, is formed by insertion of a double bond into $\Delta 9$ position of stearyl-ACP inside the plastid (Fig. 1.2). Some details of this desaturation step are described in the section 1.5.6 of this review.

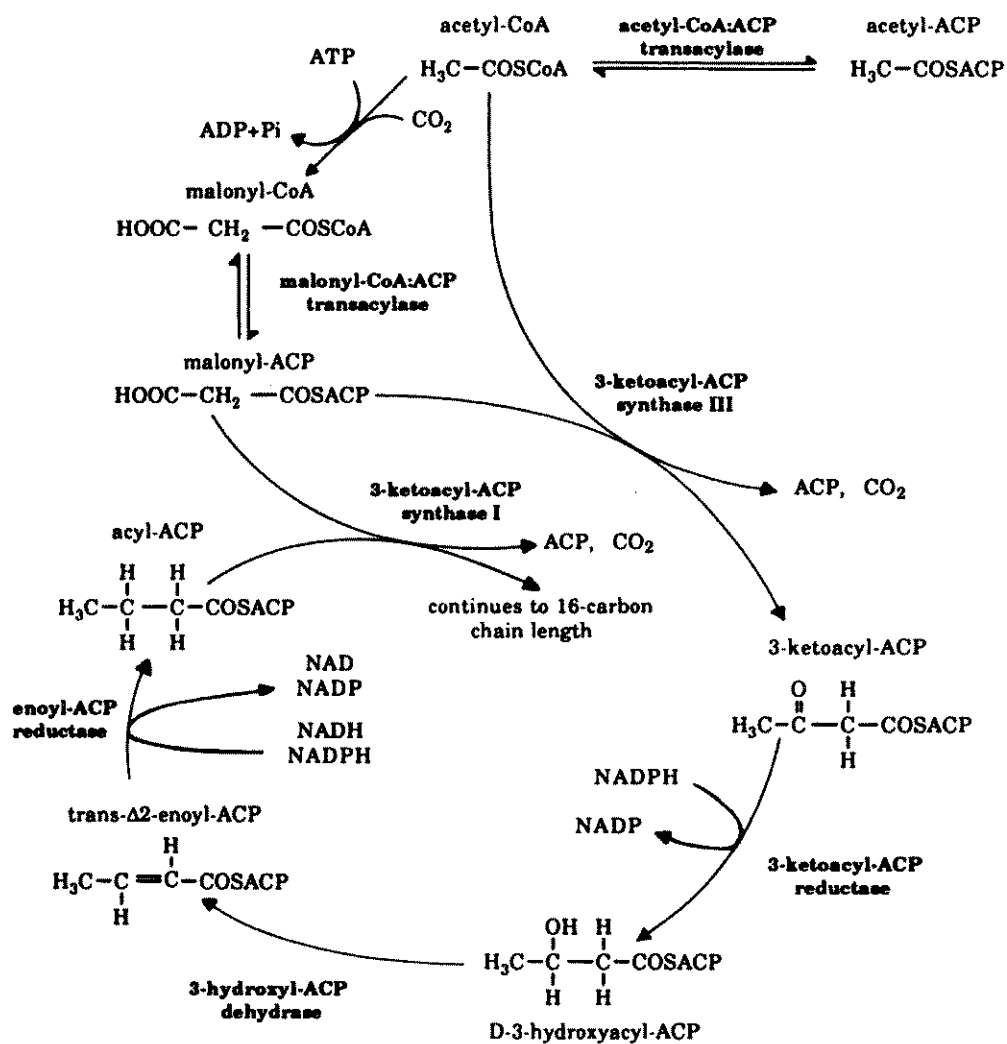


Fig. 1.1 Schematic overview of fatty acid synthesis in plants. Based on the diagram in the review article by Browse and Somerville (1995).

1.5.3 Termination of the Fatty Acid Chain and Two Parallel Pathways for Lipid Biosynthesis

The elongation of the fatty acid chain in the plastid is terminated when the acyl group is removed from ACP. As illustrated in Fig. 1.2, this can occur in two ways: either by direct use in the production of plastid membrane lipids (the 'prokaryotic' pathway), or by export to the cytoplasm as CoA esters, in which form they are then incorporated into lipids in the endoplasmic reticulum (microsome) (the 'eukaryotic' or Kennedy pathway). In the prokaryotic pathway, an acyltransferase in the plastid transfers the fatty acid from ACP to glycerol-3-phosphate (G-3-P) or to monoacylglycerol-3-phosphate. However, in the eukaryotic pathway, fatty acids are released by acyl-ACP thioesterases and are destined to pass through the plastid envelope and be subsequently converted to CoA esters. In developing oilseeds, only a small proportion of acyl groups is needed for the synthesis of plastid membrane lipids and the bulk of fatty acyl groups are exported for oil synthesis *via* eukaryotic pathway (Heinz, 1993).

The specificity of the acyl-ACP thioesterases determines the ultimate chain length of fatty acids in plant oils and the competition between the termination by thioesterase and extension by KAS enzymes determines the proportion of fatty acid with specific chain length. The seed TAG profile of California bay (*Umbellularia californica*) may consist of up to 70% lauric acid (12:0), due to the strong activity of a C12:0-ACP thioesterase. However, fatty acids of 18 and 16 carbons in length tend to predominate in most other oilseed crops. Depending on plant species, C18 fatty acids generally range from 60 to 95% and C16 fatty acids range from 5 to 45% of total fatty acids of seed storage lipids (Ohlrogge et al., 1993).

1.5.4 Extra-Plastidial Metabolism of Fatty Acids

As illustrated in Fig. 1.2, fatty acid molecules formed in the plastid stroma are released and cross the membrane to the cytosol where they are subsequently converted to fatty acyl-CoA esters through the activity of an acyl-CoA synthetase. The fatty acyl pool

serves as a source of fatty acids used in the production of phospholipids and TAG. Formation of TAG begins with the acylation of glycerol-3-phosphate at the sn-1 position. The enzyme catalysing the reaction, glycerol-3-phosphate acyltransferase (G3PAT), is located in the microsomes and is most active in acylating saturated fatty acids (Eccleston and Harwood, 1990). The resulting molecule, lysophosphatidic acid (LPA), is then acted upon by a second microsomal enzyme, the lysophosphatidic acid acyltransferase (LPAAT). The reaction forms phosphatidic acid (PA) through the transfer of a fatty acyl group to the sn-2 position. LPAAT is usually highly selective, showing the highest enzymatic activity for the acylation of unsaturated fatty acid (Ichihara, 1984; Bafor et al., 1990). PA may be converted to phosphatidylcholine (PC) by the addition of a choline headgroup or it may be dephosphorylated, forming diacylglycerol (DAG), by the activity of PA phosphatase. DAG acyltransferase (DAGAT), located in the microsomal membranes, converts DAG to TAG by acylating the sn-3 position with an acyl-CoA substrate from the fatty acyl-CoA pool. PC molecules are acted upon in the ER by membrane-bound desaturases, principally microsomal ω -6 and ω -3 desaturase. These desaturases convert the fatty acyl group on PC from oleic to linoleic acid and from linoleic to linolenic acid, respectively.

The stereospecific distribution of acyl groups in the triacylglycerols is a result of both the distribution of acyl groups in the acyl-CoA pool and specificity of the acyltransferases. Some studies which have been carried out on the acyl chain specificity of acyl transferases show that G3PAT and DAGAT are flexible in their use of acyl CoA substrates, whilst LPAAT is highly selective (Ichihara et al., 1987; Stymne and Stobart 1987). This suggests that the composition of acyl groups at positions sn-1 and sn-3 is largely determined by acyl CoA pool composition whilst sn-2 occupancy is governed by the acyl specificity of the LPAAT.

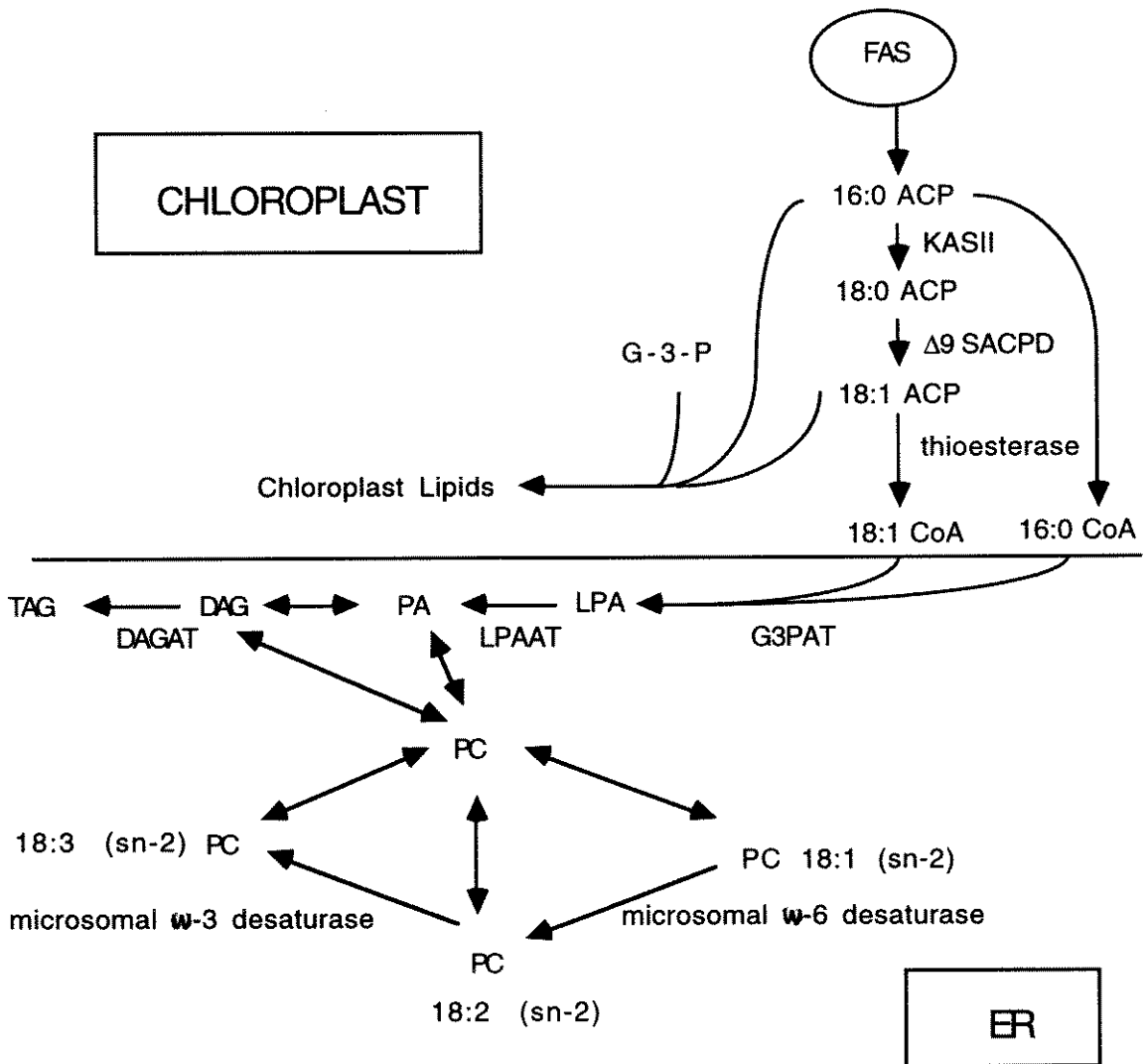


Fig. 1.2 A simplified diagram of the scheme of phospholipid and triacylglycerol synthesis in plants. KASII, β -ketoacyl-ACP synthase II; $\Delta 9$ SACPD, $\Delta 9$ stearoyl-ACP desaturase; LPA, lysophosphatidic acid; PA, phosphatidic acid; DAG, diacylglycerol; PC, phosphatidylcholine; G-3-P, glycerol-3-phosphate; G3PAT, glycerol-3-phosphate acyltransferase; LPAAT, lysophosphatidic acid acyltransferase; DAGAT, DAG acyltransferase; ER, endoplasmic reticulum.

1.5.5 Cloning of Plant FAS and Acyltransferase Genes

The basic metabolic pathways that lead to the synthesis of the major plant lipids are now mostly understood. Much of the research effort in plant lipid biosynthesis over the past half decade has been directed toward cloning genes encoding enzymes in these pathways. Most of the enzymes involved in *de novo* fatty acid synthesis have been purified and corresponding genes or cDNAs cloned. Table 1.3 provides an overview of publications on the gene cloning of those enzymes involved in plant fatty acid synthesis. This table does not include fatty acid desaturase genes which will be discussed in detail in subsequent sections in this chapter.

Table 1.3 A Summary of Gene Cloning of Plant Protein and Key FAS Enzymes Involved in the Fatty Acid Synthesis and Acyl Transfer

ENZYME	SPECIES FROM WHICH GENE HAS BEEN CLONED	ACCESSION No.	REFERENCE
ACCase	<i>Cyclotella cryptica</i>	L20784	Roessler and Ohlrogge, 1993
	<i>Arabidopsis thaliana</i>	L27074	Roesler et al., 1994
	<i>Brassica napus</i>	X77576; Y10301; Y10302.	Schulte et al., 1997; Elborough et al., 1994
	<i>Triticum aestivum</i>	Z23038; U10187.	Gornicki et al., 1994; Podkowinski et al., 1996
	<i>Medicago sativa</i>	L25042	Shorrosh et al., 1994
	<i>Zea mays</i>	U90128; A25272; A25275; A25278; A25281; A25284.	Egli et al., 1993; patent 9311243-A; Ashton et al., 1994
	<i>Glycine max</i>	L42814; L48995; L49020.	Anderson et al., unpublished

Table 1.3 con't

ACP	<i>A. thaliana</i>	X13708; X13714; X57698; X57699; L14814.	Post-Beittenmiller et al., 1989; Hlousek-Radojcic et al., 1992
	<i>Hordeum vulgare</i>	M24425; M24426; M58754; M58753; M63799.	Hansen, 1987; Hansen and Kauppinen, 1991; Hansen and von Wettstein-Knowles, 1991
	<i>B. napus</i>	X64114; X16114; X13126; X13122; X13128; X07970.	de Silva et al., 1990; de Silva et al., 1992; Safford et al., 1988
	<i>Spinacia oleracea</i>	M17636	Scherer and Knauf, 1987
	<i>Cuphea lanceolata</i>	X77620; X77621; X77622.	Voetz et al., 1994
	<i>Oryza sativa</i>	D10420	Uchimiya et al., 1992
	<i>Z. mays</i>	X57956	Souciet and Weil, 1992
	<i>Brassica campestris</i>	X70336; Y00418.	Das and Ghosh, 1995
KASI	<i>Hordeum vulgare</i>	M60410; M95172; P23902.	Siggaard-Andersen et al., 1991
	<i>Ricinus communis</i>	L13241; L13242.	Genez, 1991
	<i>C. lanceolata</i>	-	Bothmann et al., 1993
	<i>A. thaliana</i>	P52410; U24177	Millar and Kunst, 1995
KASII	<i>H. vulgare</i>	Z34268; Z34269.	Wissenbach et al., 1992
	<i>Cuphea wrightii</i>	U67316	Slabaugh et al., 1993
	<i>Simmondsia chinensis</i>	U37088	Lassner et al., 1996

Table 1.3 con't

KASIII	<i>A. thaliana</i>	Y11689; L31891; P49243.	Tai et al., 1994
	<i>S. oleracea</i>	Z22771; O07510.	Tai and Jaworski, 1993
	<i>C. wrightii</i>	U15935; U67316; P49244; P49245.	Slabaugh et al., 1995
	<i>Allium porrum</i>	U306000	Chen and Post-Beittenmiller, 1996
	<i>Porphyra umbilicalis</i>	P31176	Reith, 1993
	<i>Porphyra purpurea</i>	P51196	Reith and Munholland, 1995
3-ketoacyl-ACP reductase	<i>A. thaliana</i>	X64464; P33207.	Slabas et al., 1992
	<i>C. lanceolata</i>	P28643; X64566.	Klein et al., 1992
	<i>B. napus</i>	X64463; S60064.	Slabas et al., 1992
enoyl-ACP reductase	<i>B. napus</i>	P80030	Kater et al., 1991
acyl-ACP thioesterase	<i>Umbellularia californica</i>	M94159	Voelker et al., 1992
	<i>A. thaliana</i>	Z36910; Z36911; Z36912.	Grellet et al., 1993; Dörmann et al., 1995
	<i>B. napus</i>	X73849; X73850.	Loader et al., 1993
	<i>Brassica rapa</i>	U17098	Jones et al., unpublished
	<i>Cuphea tinctorius</i>	M96568	Knutzon et al., 1992a
	<i>Coriander sativum</i>	L20978	Dörmann et al., 1994
	<i>Cuphea hookeriana</i>	U17076	Jones et al., 1995
	<i>C. hookeriana</i> (8:0 and 10:0)	U39834	Dehesh et al., 1996
	<i>C. wrightii</i>	U56103;	Slabaugh et al., 1993
	<i>Cinnamomum camphora</i>	U31813	Yuan et al., 1995

Table 1.3 con't

FAE (fatty acid elongase)	<i>A. thaliana</i>	U29142	James et al., 1995
	<i>B. napus</i>	U50771	Roscoe et al., unpublished
glycerol-3- phosphate acyltransferase	<i>Cucumis sativus</i>	M80571	Johnson et al., unpublished
	<i>A. thaliana</i>	D00672	Nishida et al., 1993
	<i>C. tinctorius</i>	L33841	Bhella and MacKenzie, 1995
	<i>S. oleracea</i>	Z49091	Wolter, 1996
lysophosphatidic acid acyltransferase	<i>Cocos nucifera</i>	U29657	Knutzon et al., 1995
	<i>Limnanthes alba</i>	U32988	Lassner et al., 1996

1.5.6 Desaturation of Fatty Acids

A major challenge in modifying the composition of plant storage lipids is to change the degree of fatty acid desaturation. There are three major unsaturated fatty acids, found in varying levels, in most plant tissues: oleic (C18:1 Δ^9), linoleic (C18:2 $\Delta^9,12$), and α -linolenic (C18:3 $\Delta^9,12,15$) acid. They are synthesised sequentially through the activity of Δ^9 , ω -6 and ω -3 desaturase enzymes, respectively.

1.5.6.1 Δ^9 stearyl-ACP desaturase and other acyl-ACP desaturases

Acyl-ACP desaturases belong to a family of soluble enzymes which catalyse the insertion of double bonds at specific positions in ACP-bound fatty acids of defined chain lengths and are one of the major determinants of the monounsaturated fatty acid composition of vegetable oils. The most widely occurring member of this family is the Δ^9 stearyl-ACP desaturase which is associated with the synthesis of oleic acid

(C18:1^{Δ9}) in plants (Nagai and Bloch, 1968; McKeon and Stumpf, 1982; Shanklin and Somerville, 1991). Other members of this family has also recently been identified from a limited number of plant families or species, which will be briefly described in the following section.

1.5.6.1.1 $\Delta 9$ stearoyl-ACP desaturase

The stearoyl-ACP desaturase in photosynthetic tissue was first found in chloroplasts in green leaves (Nagai and Bloch, 1966, 1968). Further studies by means of ¹⁴C partitioning also located this enzyme in three other separate tissues, castor bean embryo (Nakamura and Yamada, 1974), avocado mesocarp and cauliflower bud (Weaire and Kekwick, 1975).

The soluble nature of stearoyl-ACP desaturases offered an opportunity to study a purified aerobic desaturase without first having to release the enzyme from a membrane environment. However, the purification of stearoyl-ACP desaturases initially proved difficult due to its instability and it was eventually accomplished from safflower with the introduction of a key affinity chromatography step utilising an ACP affinity column (McKeon and Stumpf, 1982). Stearoyl-ACP desaturase has now been isolated from several sources, including soybean (Kinney et al., 1990) and avocado (Shanklin and Somerville, 1991). Molecular weight determinations indicate that $\Delta 9$ stearoyl-ACP desaturase is a dimer of 68 kDa consisting of two similar units (McKeon and Stumpf, 1982). Reduced ferredoxin appears to be the immediate electron donor to this desaturase (Jaworski and Stumpf, 1974; McKeon and Stumpf, 1982). Hydrogen peroxide has been shown to be an inhibitor of the stearoyl-ACP desaturase (McKeon and Stumpf, 1982).

In 1991, two groups independently cloned the cDNAs encoding $\Delta 9$ stearoyl-ACP desaturase from castor bean (Shanklin and Somerville 1991) and safflower (Thompson et al., 1991) using antibodies raised against the purified enzymes. Since then, cDNA sequences from almost 20 plant species have become available through the GenBank database. Perhaps the most distinctive characteristic of this gene is that it encodes a

precursor protein which contains a transit peptide at the N-terminus for targeting to the plastid. In the case of safflower, the length of the transit peptide is 33 amino acids and the mature protein consists of 363 amino acids with a molecular mass of 41,000 Da (Thompson et al., 1991). Relatively high similarities, in terms of deduced amino acid sequences, have been observed among all available plant sequences, but they all have no sequence identity to either the animal or the fungal $\Delta 9$ desaturases (Shanklin and Somerville, 1991). This indicates their different evolutionary origins and reflects three major differences.

Firstly, the plant enzyme is water soluble whereas the animal and fungal enzymes are non-soluble integral membrane proteins. Secondly, the plant enzyme uses ACP whereas animal and fungal enzymes utilise CoA as the substrate carrier. Thirdly, the plant enzyme utilises ferredoxin whereas animal and fungal enzymes use cytochrome b_5 as electron donor. Interestingly, Grayburn et al. (1992) effectively decreased the saturated fatty acids content in tobacco leaves by expressing $\Delta 9$ stearoyl-CoA desaturase from rat. Similar observation was also made by expressing a yeast $\Delta 9$ desaturase gene in tobacco (Polashock et al., 1992). These indicated that the yeast and animal enzymes were correctly inserted into the endoplasmic reticulum and were able to utilise cytochrome b_5 as an electron donor. Normally the free saturated fatty acids, such as 16:0 and 18:0, will be converted to CoA after being exported from the plastid and then incorporated into glycerol lipid without further desaturation. However, the presence of the $\Delta 9$ stearoyl-CoA desaturase introduced a novel pathway of desaturation in endoplasmic reticulum.

Comparison of the deduced amino acids of the $\Delta 9$ desaturase gene from rat, yeast and plants revealed a conserved pair of the amino acid sequence (D/E-E-X-R-H) separated by about 100 amino acids (Fox et al., 1993). These motifs have been suggested to produce a powerful mechanism for oxygen activation and can therefore be used to catalyse energetically demanding reactions, such as the abstraction of hydrogen to form double bonds in fatty acyl or other inactivated hydrocarbon chains. Studies with

purified castor bean enzyme demonstrated that it contained four atoms of iron per homodimer and these iron ions comprise a diiron-oxo cluster (Fox et al., 1993). This is consistent with it belonging to a class of O₂ activating proteins containing diiron-oxo clusters. Other enzymes of this class are ribonucleotide reductase (Nordlund et al., 1993) and methane monooxygenase (Rosenzweig, et al., 1993). The recent available crystal structure of the castor $\Delta 9$ stearoyl-ACP desaturase shows the diiron structure within a four-helix bundle of the protein (Lindqvist et al., 1996). The iron ions are bound in a highly symmetric environment and the iron centre is buried in the interior of the $\Delta 9$ stearoyl-ACP desaturase. A deep channel lined with hydrophobic residues connects the surface of the enzyme to the active site and neatly accommodates the 18-carbon substrate. When the 18:0 acyl chain substrate is present in the channel, the hydrogens on carbon nine and ten are presented to the active site. It is presumed that during catalysis, a peroxide radical could be generated, and be capable of abstracting one of the hydrogen atoms at the C9 position of the fatty acid.

1.5.6.1.2 Other acyl-ACP desaturases

Besides the $\Delta 9$ stearoyl-ACP desaturase, at least four other different acyl-ACP desaturase genes have been isolated so far, these include a $\Delta 4$ palmitoyl-ACP desaturase of *Corianderum sativum* seed (Cahoon et al., 1992; Cahoon and Ohlrogge, 1994), a $\Delta 6$ palmitoyl-ACP desaturase of *Thunbergia alata* seed (Cahoon et al., 1994), a $\Delta 9$ myristoyl -ACP desaturase of *Pelargonium xhortorum* trichomes (Schultz et al., 1996) and a cDNA from *Asclepias syriaca* encoding a $\Delta 9$ acyl-ACP desaturase which exhibited 7-10 fold greater specificity for $\Delta 9$ palmitoyl-ACP and 30-fold greater specificity for $\Delta 9$ myristoyl-ACP than did known $\Delta 9$ stearoyl-ACP desaturases (Cahoon et al., 1997b). As their names indicate, these enzymes introduce double bonds into different positions of fatty acids of different chain lengths. In spite of their different functional properties, these enzymes share more than 70% amino acid sequence similarity. The most notable difference between the primary structures of the $\Delta 9$

stearoyl-ACP desaturase and the four diverged acyl-ACP desaturases is the variation in the number of amino acids in a region near their N-termini.

The unveiling of the architecture of the substrate channel with respect to the diiron centre in castor stearoyl-ACP desaturase is a critical step towards an understanding of substrate specificity and positional specificity of the acyl-ACP desaturase family (Lindqvist et al., 1996). This information will be useful for the rational redesign of desaturase enzymes with desired chain length and regio-specificities by altering the geometry of the substrate channel. In fact, a recent study with site-directed mutagenesis has demonstrated that on the basis of the active-site model, the substrate and regio-specificities of acyl-ACP desaturases can be modified by the replacement of specific amino acid residues (Cahoon et al., 1997b). For instance, a $\Delta 9$ stearoyl-ACP desaturase was converted into an enzyme with substrate preference for $\Delta 9$ palmitoyl-ACP by replacing of two amino acid residues. Similarly, the $\Delta 6$ palmitoyl-ACP desaturase was converted into an enzyme that functions principally as a $\Delta 9$ stearoyl-ACP desaturase by replacement of five amino acids. These results, along with additional information from crystallographic studies of acyl-ACP desaturases will lead to the design of enzymes for the production of novel monounsaturated fatty acids in transgenic oilseed crops.

1.5.6.2 Membrane-bound desaturases

Unlike the acyl-ACP desaturases, both ω -6 and ω -3 desaturases are membrane-bound enzymes (Gurr et al., 1969; Nichols et al., 1967; Nichols, 1969) and use the fatty acid at the sn-2 position of phosphatidylcholine (PC) as substrates (Stymne and Appelqvist, 1978; Stymne et al., 1980; Harwood, 1988; Jaworski et al., 1987). The desaturation reactions catalysed by these membrane-bound desaturases also require molecular oxygen and the primary reductant, NAD(P)H. Electron transport, between NAD(P)H and the desaturase enzyme, is mediated through cytochrome b_5 and NADH:cytochrome b_5 reductase (Smith et al., 1990).

However, biochemical characterisation of the membrane desaturases by traditional approaches has been limited because they are difficult to purify (Schmidt et al., 1994). Instead, much of the understanding of the reactions and regulation of membrane desaturation has come from the isolation and characterisation of seven classes of *A. thaliana* mutants, each one deficient in a particular desaturation step (James and Dooner, 1990; Lemineux et al., 1990; Browse and Somerville, 1995). These mutants have enabled the isolation of the corresponding genes by map-based cloning or gene tagging (Arondel et al., 1992; Okuley et al., 1994; Yadav et al., 1993).

1.5.6.2.1 Cloning of ω -6 desaturase genes

Mutants of *A. thaliana* at the fatty acid desaturation 2 (*fad2*) locus are deficient in activity of the microsomal ω -6 desaturase responsible for production of polyunsaturated lipids in the eukaryotic pathway of lipid synthesis (Miquel and Browse, 1992), while the *fad6* mutant is deficient in activity of plastidial ω -6 desaturase responsible for the synthesis of polyunsaturated fatty acids by the prokaryotic pathway. A *FAD2* allele was identified in a population of *A. thaliana* in which mutation had been created by T-DNA insertions (Okuley et al., 1994). Subsequently, the genomic DNA flanking the T-DNA was cloned and used to isolate cDNA and genomic clones. When a cDNA containing the entire *FAD2* coding region was expressed in *fad2* mutant plants it was shown that the mutant fatty acid phenotype was complemented (Okuley et al., 1994). The microsomal ω -6 desaturase gene has been isolated more recently from *B. juncea* (Singh et al., 1995) and soybean (Heppard et al., 1996).

A putative *FAD6* cDNA clone was also isolated from *A. thaliana* using a PCR based approach and shown to restore normal levels of fatty acids in transgenic plants of a *fad6* mutant (Falcone et al., 1994). At the same time as the *A. thaliana* *FAD6* gene was isolated, a plastidial ω -6 desaturase from spinach (*Spinacia oleracea*) was purified from chloroplast envelop membranes by three simple chromatography steps (Schmidt et al., 1994). The N-terminal amino acid sequence of protein was determined and used for

the construction of degenerate PCR primers and subsequently a full length cDNA encoding the plastidial ω -6 desaturase from spinach was cloned (Schmidt et al., 1994). The plant plastidial ω -6 desaturase cDNAs has a high degree of sequence homology to the ω -6 desaturase of cyanobacterium, which also uses 'chloroplast' lipid substrates. By contrast, they show poor homology to any other known higher plant desaturase.

1.5.6.2.2 Cloning of ω -3 desaturase genes

By characterisation of the *A. thaliana* mutants that are deficient in the activity of the ω -3 desaturases, three different loci, *FAD3*, *FAD7* and *FAD8* were found to be responsible for the production of trienoic fatty acids (McConn and Browse, 1996). The cDNAs corresponding to these loci have been isolated from *A. thaliana*: the *FAD3* gene encodes a microsomal ω -3 desaturase (Arondel et al., 1992; Yadav et al., 1993), while the *FAD7* (Iba et al., 1993) and *FAD8* (Gibson et al., 1994) gene encode two different plastid ω -3 desaturases. The *FAD8* gene is induced by low temperature, whereas expression of the *FAD7* gene is detectable at both normal and low temperatures. The genes for the microsomal ω -3 desaturase have also been isolated from a variety of other plant species, including *Glycine max* (Yadav et al., 1993), *Brassica napus* (Yadav et al., 1993), *Vigna radiata* (Yamamoto, 1994), *Nicotiana tabacum* (Hamada et al., 1994) and *Oryza sativa* (Kodama et al., 1997). In addition, genes for plastidial ω -3 desaturase were isolated from *G. max* (Yadav et al., 1993), *B. napus* (Yadav et al., 1993), *R. communis* (van der Loo and Somerville, 1994), *Limnanthes douglasii* (Bhalla and Mackenzie, 1995) and *Petroselinum crispum* (Kirsch et al., 1997b). While relatively low sequence identity is observed between plastidial and microsomal ω -6 desaturases in plants, a high degree of sequence identity between the two types of ω -3 desaturases exists. This indicates that they share a recent common ancestor in their evolutionary history. More recently, the first animal ω -3 fatty acid desaturase has been isolated from *Caenorhabditis elegans* (Spychalla et al., 1997). It shows equal homology to both the microsomal ω -6 desaturase encoded by *FAD2* and microsomal ω -3 desaturase encoded by *FAD3* in *A. thaliana*. When expressed in transgenic plants, it

resulted in a 90% increase in the proportion of α -linolenic acid in root lipids (Spychalla et al., 1997).

1.5.6.2.3 Cloning of other membrane-bound desaturase genes

Gamma-linolenic acid (GLA; C18:3 $\Delta^{6,9,12}$) is a component of the seed oils of evening primrose (*Oenothera* spp.), borage (*Borago officinalis* L.), and some other plants. It is widely used as a dietary supplement and for treatment of various medical conditions. GLA is synthesised by a Δ^6 fatty acid desaturase using linoleic acid as a substrate. The Δ^6 fatty acid desaturase in the developing cotyledons of borage is similar to the ω -6 and ω -3 desaturases in its location and substrate specificity, and is assumed to use cytochrome b_5 as its electron donor (Griffiths et al., 1989; Smith et al., 1990). A cDNA corresponding to the gene encoding for the Δ^6 desaturase was isolated from borage using a PCR approach with degenerate primers based on the sequences of the first and third histidine boxes in the membrane-bound desaturases (Sayanova et al., 1997). It has reasonable sequence identities with other plant microsomal ω -6 and ω -3 desaturases, but it is also characterised by the presence of an N-terminal extension related to the cytochrome b_5 class of haem-binding proteins. Interestingly, this domain is not present in the plant microsomal ω -6, ω -3 desaturases and ω -6 hydroxylase from castor bean, although the use of microsomal cytochrome b_5 as an electron donor by these enzymes has been clearly demonstrated (Smith et al., 1990; 1994).

Fukuchi-Mizutani et al. (1995) isolated three cDNA clones for senescence-inducible genes from a cDNA library derived from mRNAs from rose petals. The amino acid sequence deduced from these cDNAs is highly homologous to those of Δ^9 acyl-lipid desaturases of cyanobacteria (Sakamoto et al., 1994) and, to a lesser extent, to the Δ^9 acyl-CoA desaturases of a yeast and mammals (Thiede et al., 1986; Stukey et al., 1990). Northern blot analysis indicated that the transcripts of the cDNAs were expressed specifically in petals at late developmental stages and during senescence. It is proposed

that such a $\Delta 9$ desaturase in the senescencing petals plays an important role in the degradation of saturated fatty acids of membrane lipids (Fukuchi-Mizutani et al., 1995).

1.5.6.2.4 Conserved domains in membrane-bound desaturases

Examination of the deduced amino acid sequences for all the membrane desaturases isolated from mammals, fungi, insects, higher plants and cyanobacteria has revealed three regions of conserved sequence containing HX(3 or 4)H, HX(2 or 3)HH, and HX(2 or 3)HH (Fox et al., 1994). These motifs are also present in the bacterial membrane enzymes alkane hydroxylase (ω -hydroxylase) and xylene mono-oxygenase. By site-directed mutagenesis, eight conserved His residues in the rat $\Delta 9$ stearyl-CoA desaturase (also a membrane bound desaturase) were individually converted to Ala; and it was demonstrated that each His to Ala mutation failed to complement the yeast *Ole1* mutant (Shanklin et al., 1994). Therefore, the conserved His residues are essential for catalytic function. In addition, spectroscopic evidence indicates that the membrane-bound desaturases are also diiron-oxo proteins (Shanklin et al., 1994). One possible role for these His residues would be to act as ligands for the iron atoms contained in these enzymes. Furthermore, all the plant fatty acid desaturases studied so far, both soluble and membrane-bound, also have several important similarities including the catalytic requirement for iron (Nagai and Bloch, 1968) and the inhibition by metal chelators (Jaworski and Stumpf, 1974). These similarities suggest a common mechanism is used for the desaturation reaction, which may include the requirement for a structurally similar catalytic site.

1.6 PROGRESS IN GENETIC ENGINEERING OF FATTY ACID COMPOSITION IN OILSEED CROPS

In principal, there are two approaches for manipulating lipid biosynthesis through genetic engineering; either a gene for a new enzyme with a novel specificity can be

introduced; or genetically the expression of an existing gene can be suppressed or enhanced.

1.6.1 Introduction of Novel Genes

With the ability to introduce novel enzymes into oilseed plants, plant genetic engineering can transcend the limits of conventional breeding. For example, the first commercially marketed 'designer oil' for human consumption, high laurate canola, was produced by the introduction of an novel lauroyl-ACP thioesterase gene from the California Bay tree into canola (Voelker et al., 1992). Lauroyl-ACP thioesterase causes premature chain termination of the growing fatty acid resulting in an accumulation of C12, rather than the C18 fatty acids that normally occur in canola oil.

1.6.2 Downregulation of Unwanted Traits

The two methods most commonly used to suppress expression of endogenous genes are 'antisense' and 'cosuppression'. A gene oriented with the N-terminal translating region immediately following the promoter in a gene construct is said to be of the 'sense orientation' while the reverse orientation, with C-terminal immediately following the promoter is called 'antisense'. When inserted into a plant genome, the sense gene construct normally produces the enzymatic activity encoded by the gene. If the gene is native to the plant, the introduction of additional gene copies may raise the level of the native enzyme beyond standard levels. At times the use of sense gene constructs had the unexpected effect of decreasing the level of homologous enzymes. This phenomenon is termed 'cosuppression' or 'trans-switching' (Heinkoff and Dressen, 1989). If an antisense gene construct is inserted into the genome, the effect may be to reduce the levels of endogenous enzymes coded by genes highly homologous to the introduced gene (Resenberg et al., 1985; van der Krol et al., 1988).

The mechanisms for cosuppression- or antisense-mediated downregulation of gene expression are poorly understood at present. One of the several proposed mechanisms of cosuppression is a biochemical switch model, where extreme sensitivity to threshold concentrations of gene products results in a switch to a new state of lower gene expression (Matzke and Matzke, 1995; Meyer and Saedler, 1996). In the case of antisense, it is generally believed that when antisense gene is introduced and transcribed, it will produce mRNA molecules that are complementary to the endogenous sense mRNA. The sense and antisense mRNA strands hybridise, forming double-stranded molecules that are blocked from leaving the nucleus and are rapidly degraded (Bourque, 1995).

The first successful alteration of fatty acid metabolism *via* genetic engineering was reported by Knutzon and co-workers (1992b) using the antisense suppression strategy to reduce levels of the stearoyl-ACP desaturase in canola. The decrease in stearoyl-ACP desaturase activity resulted in a build up of the stearoyl-ACP substrate which allowed more stearate to be released from ACP *via* the activity of stearoyl-ACP thioesterase. As a result, the content of stearic acid in the engineered rapeseed oil was increased up to 20-fold (40%).

Since the recent availability of cDNA sequences encoding for membrane bound fatty acid desaturases, notable progress has been made in the production of high oleic soybean and canola oils by the suppression of the ω -6 fatty acid desaturase gene *via* both antisense and co-suppression. The oleate level in canola was reported to be increased up to 83% in comparison with 63% in the control (Murphy, 1996). In soybean, antisense engineering produced a line containing 79% oleate, while up to 88% oleate content were found in transgenic soybean lines transformed with a cosuppression construct (Kinney, 1995).

1.7 SYNOPSIS OF THE WORK EMBODIED IN THIS THESIS

The cotton industry has allocated much of its resources in the past to the more urgent issue of pest control and fibre quality traits, and the modification of cottonseed oil has been a relatively neglected area of research. As a result, most of the enzymes and their encoding genes involved in lipid metabolism in cottonseed have never been isolated and the temporal and spatial controls over gene expression during fatty acid synthesis and oil accumulation are largely unknown. The recent progress in the understanding of oil biochemistry and genetic engineering, as summarised above, has demonstrated the feasibility of producing novel or high value seed oils in a wide range of oilseed crops. To fully realise the potential benefits of applying these techniques to cottonseed oil the key genes involved in the fatty acid synthesis and modification will need to be isolated and characterised. The present research is concerned with producing high quality cottonseed oil for human consumption. The primary goals include the production of high oleic oil for deep frying, high stearic oil for use as spreadable fat, such as margarine etc and cocoa butter substitute. Linoleic acid is the major fatty acid and the only polyunsaturated fatty acid in cottonseed. There are, therefore, two major fatty acid desaturation steps in the synthesis of cottonseed oil; They are catalysed by the plastidial Δ^9 stearoyl-ACP desaturase and microsomal ω -6 desaturase. The present project focused on the isolation and characterisation of the genes encoding for these two key desaturase enzymes from cotton.

Following this literature review, the construction of a cDNA library from developing cotton embryos and the general materials and methods used in this study are described in Chapter Two. A survey of lipid content and fatty acid composition in a small collection of cotton varieties and a few closely related *Gossypium* species, along with an analysis of the fatty acid profile in developing cottonseed embryos is presented in Chapter Three. Chapters Four and Five describe molecular cloning and characterisation of Δ^9 stearoyl-ACP desaturase and microsomal ω -6 desaturase from cotton, respectively. The genomic structures of these genes are also described. A large 5' UTR

intron was isolated from the seed-specific microsomal ω -6 desaturase gene, pghD12-1, from *G. hirsutum* and 23 other *Gossypium* species. On the basis of the DNA sequence of this intron, several evolutionary aspects of the pghD12-1 gene, including genome introgression and gene redundancy in tetraploids is discussed in Chapter Six. Chapter Seven is a general discussion on the subject of modifying cottonseed oil quality by genetic engineering.

Chapter Two

General Materials and Methods

The experiments described in this thesis were carried out in the Division of Plant Industry, CSIRO Australia and Plant Breeding Institute, Cobbitty, Sydney University. All chemicals used for *in vitro* use were at least analytical grade in standard. Solutions were prepared under sterile conditions using MilliQ H₂O, and autoclaved when appropriate. All the restriction enzymes used were purchased from New England Biolabs. All the polymerase chain reactions (PCR) were performed using AmpliTaq® DNA polymerase (Perkin Elmer, USA). The origin of other enzymes, kits used are indicated together with suppliers' names in the following sections of this chapter and in the remaining chapters. The synthetic oligonucleotides were made on an Applied Biosystems (USA) Model 377 DNA synthesiser by Lynda Graf in the Division of Plant Industry, CSIRO Australia. α -³²P[dCTP] was supplied by Bresatec, Australia. The DNA sequencing reactions were made using ABI PRISM™ kits (Perkin-Elmer). The sequence gel electrophoresis was carried out using an ABI373 DNA sequencer by Gina Leach in the Division of Plant Industry, CSIRO Australia. The general materials and methods routinely used throughout this study are described below. The materials and methods that were used only in particular parts of this study are described in the individual chapters.

2.1 GROWTH OF PLANT MATERIALS

Potting soils were prepared by the plant growth facility at the Division of Plant Industry, CSIRO Australia. Unless otherwise stated, all the cotton plants were grown on a soil mix consisting of 75% Pryor's mix and 25% Potting mix. All the cotton plants were grown in the glasshouse at 28/17°C day/night, with 16 hours photoperiod.

2.2 GROWTH OF BACTERIA

Unless otherwise indicated, cultures of *E. coli* bacteria were grown overnight at 37°C, using solid or liquid Luria Broth media (LB; 10 g/l Bacto-tryptone, 5 g/l Bacto-yeast extract, 5 g/l NaCl, pH7.0). Each liquid culture was inoculated using a single bacterial colony and grown on a rotator. The appropriate antibiotics, either Ampicillin (100 µg/ml) or Kanamycin (50 µg/ml) were added to bacterial growth media at the final concentration of 100 µg/ml and 50 µg/ml, respectively.

2.3 TRANSFORMATION OF *E. COLI* WITH PLASMIDS BY ELECTROPORATION

Transformation of competent cells with plasmids was performed according to the recommendations supplied with the Gene-Pulser (BioRad). Competent cells (50 µl) were combined with 5 µl deionised H₂O (dH₂O) containing 5 ng plasmid DNA or 60 ng DNA from a ligation reaction. This mixture was transferred to a pre-chilled disposable Gene-Pulser[®]/E.coli Pulser[™] cuvette, 0.2 cm (BioRad), and subject to electroporation using a Gene-Pulser (BioRad), set at 1.8 kV, 125 µFD and 200 w. Immediately following electroporation, the cells were mixed with 600 µl LB broth without antibiotic, transferred into a new eppendorf tube and incubated at 37°C for half a hour. 100 µl of culture were then plated onto a plate of solid LB media supplemented with appropriate antibiotic, and grown at 37°C overnight.

2.4 DNA AND RNA ISOLATION AND PURIFICATION

2.4.1 Plasmid DNA Miniprep

A single bacterial colony was transferred into 2 ml of LB medium containing the appropriate antibiotic in a loosely capped 15-ml tube. The culture was incubated overnight at 37°C with vigorous shaking. 1.5 ml of the culture was poured into an eppendorf tube and centrifuged at 12,000x g for 1 min at RT in a microfuge. The medium was removed by aspiration, leaving the bacterial pellet as dry as possible. 200 µl lysis buffer (0.1M NaOH and 0.5% SDS) was added and mixed by pipetting up and down. 200 µl 3M NaAc pH6.0 was added and mixed by inverting many times and incubated on ice for 5 min. Following centrifugation for 10 min at room temperature, the supernatant was transferred into a new tube. Nucleic acids was precipitated by adding 1 ml cold ethanol, vortex to mix and centrifuge for 10 min at 4°C. The pellet was washed in 70% ethanol by vortexing, centrifuging for 5 min at room temperature. After a briefly drying in a speedvac, the pellet was resuspended in 30 µl dH₂O.

2.4.2 Purification of Plasmid DNA for DNA Sequence Determination

There are two key features about this modified alkaline-lysis plasmid DNA miniprep procedure. Firstly, bacteria were grown in Terrific Broth (TB), instead of LB; this results in a four- to eight-fold increase in the number of bacteria per millilitre of media, which leads, in turn, to higher plasmid yield. Secondly, the precipitation of plasmid DNA by PEG₈₀₀₀ (polyethylene glycol) yields high-quality, super-coiled plasmid DNA that is relatively free of contaminating chromosomal DNA and RNA. One litre of TB medium was made by mixing 100 ml of a sterile solution of 0.17 M KH₂PO₄, 0.72 M K₂HPO₄ and 900 ml of sterile, cool solution containing 12 g Bacto-tryptone, 24 g Bacto-yeast extract and 4 ml glycerol.

A single colony was transferred to 2 ml of TB medium and incubated overnight at 37°C, with an appropriate amount of antibiotic, in 50-ml polypropylene tubes. 1.5 ml

aliquots of culture pelleted by centrifugation for 1 min in a microcentrifuge. The supernatant was removed and the bacterial pellet was resuspended in 200 μ l of GTE buffer (50 mM glucose, 25 mM Tris-HCl, 10 mM Na₂EDTA, pH 8.0) by pipetting up and down. 300 μ l of freshly prepared lysis buffer (0.2N NaOH and 0.1% SDS) were added and mixed by inversion and incubated on ice for 5 min. This solution was neutralised by adding 300 μ l of 3.0 M potassium acetate (pH4.8) and mixed by inverting the tube, and incubated on ice for 5 min. The cellular debris was removed by centrifugation for 10 min at room temperature, and the supernatant was transferred to a clean tube. RNaseA (Sigma, USA) was added to a final concentration of 20 μ g/ml and incubated at 37°C for 20 min. The supernatant was extracted twice with equal volume of phenol/chloroform/isoamyl alcohol (25:24:1). The plasmid DNA was precipitated by adding an equal volume of 100% isopropanol and pelleted by centrifugation for 10 min at room temperature. The DNA pellet was washed with 700 μ l of 70% ethanol and then dried in a speedvac. The pellet was dissolved in 32 μ l of dH₂O, and then precipitated by adding 8.0 μ l of 4 M NaCl and 40 μ l of autoclaved 13% PEG₈₀₀₀. After the thorough mixing, the sample was incubated on ice for 20 min, and then the plasmid DNA was pelleted by centrifugation for 15 min at 4°C. The supernatant was carefully removed and the pellet was rinsed with 700 μ l of 70% ethanol. The DNA pellet was dried in a speedvac and then resuspended in 20 μ l of dH₂O.

2.4.3 Isolation of Genomic DNA from Plant Materials

All the genomic DNA samples used in this studied was prepared according to Paterson et al. (1993). Three grams of leaf tissues were ground in liquid nitrogen to fine powder and kept in -20°C freezer until all the samples were ground. The powder was transferred to 20 ml of cold nuclear buffer and inverted to homogenise. The nuclear buffer contains 1.0 M glucose, 100 mM Tris-HCl, 5 mM Na₂EDTA and 2% PVP-40 (polyvinylpolypyrrolidone, Mw 40,000), 0.1% DIECA (diethyldithiocarbamic acid), 0.1% ascorbic acid, 0.2% mercaptoethanol, pH7.5. Following centrifugation at 1000x g

for 20 min at 4°C, the pellets were suspended in 10 ml of 65°C lysis buffer and incubated for 20 min in 65°C water bath. The lysis buffer contains 100 mM Tris-HCl pH8.0, 1.4 M NaCl, 20 mM Na₂EDTA, 2% CTAB (hexadecyl triammonium bromide), 2% PVP-40, 0.1% DIECA, 0.1% ascorbic acid and 0.2% mercaptoethanol. Equal volume of chloroform/isoamyl alcohol was added and mixed by inverting many times. Following centrifugation at 4,500x g for 10 min, the upper aqueous phase was transferred to a new tube. This extraction step was repeated once more prior to the precipitation of DNA by addition of two volumes of 100% ethanol. The DNA was hooked out with a glass hooker and rinsed in 70% ethanol, air dried and resuspended in 500 µl of TE buffer. This was followed by further two successive extractions with equal volume of phenol/chloroform/isoamyl alcohol and the DNA was then precipitated by adding 1/10 volume of 3M NaAc and two volume of 100% ethanol. Following rinsing in 70% ethanol, the DNA was air dried and resuspended in 1 ml of 1M CsCl in TE buffer.

2.4.4 Purification of Genomic DNA by CsCl Gradient

2.5 ml of 5.7 M CsCl in TE was added to a Beckman polyallomer 13x51 mm tube (non sealing) using a disposable pipette. 30 µl EtBr (10 mg/ml) was added to the plant genomic DNA in 1.0 M CsCl-TE buffer and laid on the top of the CsCl cushion. Following centrifugation at 35,000x rpm for 16 hours at 18°C in a SW50.1 rotor, the DNA band was removed using a 1-ml syringe and 18 gauge needle and added to a new 10-ml tube containing 5 ml of CsCl-saturated isopropanol. After gentle mix and the top layer containing EtBr was discarded. Three more extractions with 1 ml of CsCl-saturated isopropanol were carried out. DNA was hooked out with a glass rod following adding three volume of 70% ethanol. After resuspended in 500 µl of TE, the DNA was precipitated by adding 50 µl of 3 M NaAc and 1 ml of 100% ethanol and rinsed in 70% ethanol prior to resuspension in 300 µl TE.

2.4.5 Isolation of RNA

Sterile, disposable plasticware which is essentially free of RNases was used for the preparation and storage of RNA without pretreatment. Glassware used for RNA isolation were baked at 180°C overnight. All the solutions were treated with 0.1% DEPC (diethyl pyrocarbonate) for at least 12 hours at 37°C and then autoclaved for 20 min at 15 ib/sq.in. on liquid cycle.

Two grams of cotton embryos or leaf tissues were ground in liquid nitrogen and mortared to a fine powder which was subsequently prickled into a beaker containing 22 mls cold extraction buffer and stirred constantly. The extraction buffer consists of 200 mM Tris-HCl pH8.5, 1.5% Lithium dodecylsulfate, 300 mM LiCl, 10 mM Na₂EDTA, 1% sodium deoxycholate, 1% Nonadet P-40. This was followed by adding 5% insoluble PVP, 90 mM mercaptoethanol, 10 mM DTT (dithiothreitol), 0.1% DEPC and stirred for 10 min prior to being transferred to a Corex tube. Then 18.4 ml of 3M ammonium acetate was added and mixed well. It was centrifuged at 6,000x rpm for 20 min at 4°C. The supernatant was transferred to a new tube and precipitated by adding 1/10 volume of 3 M NaAc, pH5.2 and 1/2 final volume of cold isopropanol and stored at -20°C for 1 hour prior to centrifugation at 6,000x rpm for 30 min using a swing rotor. The pellet was resuspended in 1 ml dH₂O and transferred to two eppendorf tubes (500 µl in each tube). The suspension was extracted with an equal volume of phenol/chloroform/isoamyl alcohol solution (25:24:1) and the phases were separated by centrifugation for 5 min at 4°C. The aqueous top layer was carefully transferred into a new eppendorf tube and it was extracted again with chloroform as above. Half volume of 5M LiCl was added to the aqueous sample, mixed well and left on ice for 3 hours prior to centrifugation at 12,000x rpm for 15 min at 4°C. The pellet was resuspended in 50 µl dH₂O. Finally, the RNA sample was precipitated by adding 5 µl NaAc and 138 µl cold ethanol and incubated on dry ice for 30 minute prior to centrifugation for 15 min at 4°C. The RNA pellet was dried in a speedvac and then dissolved in 30 µl RNase-free H₂O.

2.5 DNA AND RNA GEL ELECTROPHORESIS AND BLOTTING

2.5.1 Preparation of a DNA-Minigel

Agarose mini-gel was cast from 50 ml of 1% melted agarose solution containing 1x TAE buffer (0.04 Tris-acetate, 1.0 mM Na₂EDTA, pH 8.0), using a 5x 8 cm mould and comb for making wells of 15 µl volume. DNA samples were mixed with 0.2 volume of 6x FLB loading buffer (15% ficoll 400, 0.25% bromophenol blue, 0.25% xylene cyanol) and electrophoresed at 50-70 V in 1x TAE buffer. DNA size marker of EcoRI cut SPP-1 DNA (Bresatec) was run alongside the sample DNAs. DNA Gel was soaked in 0.5 mg/l solution of ethidium bromide for 10 min, destained by rinsing in water for 10 in and photographed using UV light of wavelength 302 nm.

2.5.2 DNA Southern Transfer

5.0 µl of cotton genomic DNA (approximately 10.0 µg DNA) was digested at 37°C overnight in a 20-µl reaction containing 1x concentration of the buffer supplied with the restriction enzyme. An 0.7% agarose gel in 1x TBE (0.045 M Tris-borate, 1.0 mM Na₂EDTA, pH8.0) was cast using a 15x 20 cm mould. Genomic DNA digests were electrophoresed on the agarose gel overnight at 20 mA in 1x TBE buffer, until the bromophenol blue from the loading dye had run 3/4 of the length of the gel. Each gel was shaken gently in 500 ml of 0.1 M HCl for 5 min, followed by rinsing and shaken in dH₂O for a further 5 min to remove the acid. DNA was transferred from the gel to Hybond N⁺ membrane (Amersham) by Southern transfer (Southern, 1975) for 4 hours, using 0.4 N NaOH as the transfer solution. After transferring, the membrane was rinsed in 2x SSC (1.75% NaCl, 0.88% sodium citrate, pH7.0) briefly and blotted dry between 2 sheets of 3M filter paper at room temperature.

2.5.3 RNA Northern Transfer

To prepare a 300 ml of 1% agarose gel for RNA electrophoresis, 3 grams agarose and 30 ml 10x MOPS [3-(N-morpholino)propanesulfonic acid] buffer were added to 255 ml dH₂O and dissolved by heating in a microwave. 16.2 ml of 37% formaldehyde were added when it cooled to 50°C and mixed by swirling. 20 µg sample of total RNA was mixed with 20 µl of loading buffer and it was heated to 95°C for 2 min to denature RNA. The loading buffer was freshly prepared by adding the following: 0.72 ml formamide, 0.16 ml 10x MOPS buffer, 0.26 ml formaldehyde (37%), 0.1 ml 80% glycerol, 0.08 ml bromophenol blue (saturated solution) and 0.18 ml dH₂O. The RNA samples were electrophoresed in 1x MOPS buffer. 1 µl ethidium bromide was also added with each sample. The electrophoresis was stopped when bromophenol blue dye migrated three-fourths down the gel. The gel was then photographed on a UV transilluminator before blotting. The two abundant RNA species, 28S rRNA (5 kb) and 18S rRNA (2 kb) were used as molecular weight standards. The RNA gel was rinsed for 20 min each in two changes of 500 ml of 10x SSC to remove formaldehyde from gel. This was immediately followed by transferring the RNA to Hybond N⁺ membrane by 0.05 N NaOH for 3 hours.

2.6 PREPARATION OF ³²P-LABELLED DNA PROBES

Radioactively labelled probes were synthesised by random priming, essentially as described by Feinberg and Vogelstein (1983). A kit for random priming was purchased from Amersham (USA). Purified cloned insert DNA (20 ng) was combined with 6.0 µl of random oligonucleotides (provided by the manufacturer). This the mixture was incubated at 95°C for 5 min to denature the DNA and quickly chilled on ice for a five min. 10 µl probe labelling buffer (0.5 M HEPES, 0.125 M Tris-HCl, 12.5 mM DTT, 12.0 mM MgCl₂, 1.0 mg/ml BSA), 2.5 µl dNTP mixture (0.2 mM dATP, 0.2 mM dTTP, 0.2 mM dGTP), 1.0 u Klenow fragment, 3.0 µl α-³²P-dCTP (10 µCi/µl), and enough MilliQ H₂O to make the total volume 25 µl. Probe labelling reactions were

incubated at 37°C for 1 hour. To remove the unincorporated dNTPs, the probe was purified by going through a Sephadex[®] G-50 NICK[®] Column (Pharmacia) according to the manufacturer's instruction.

2.7 HYBRIDISATION AND AUTORADIOGRAPHY

The prehybridisation was carried out by placing membranes in a hybridisation bottle containing 5-20 ml of prehybridisation solution in a hybridisation oven equipped with a rolling apparatus for at least 4 hours at 42°C. The prehybridisation/hybridisation buffer was based on the recipe described by Khandjian (1987). It contains 50 mM Tris-HCl pH7.5, 1M NaCl, 50% formamide, 10x Denhardt's solution, 10% dextran sulfate, 1% SDS, 0.1% sodium pyrophosphate, 0.1 mg/ml herring sperm DNA.

Following prehybridisation, probe was denatured by heating for 10 min at 95°C, cooled on ice for five min before it was added to the hybridisation mixture. The hybridisation was performed at 42°C for 16 hours. Unless otherwise stated, membranes were then briefly washed in 2x SSC, 0.1% SDS at 65°C. This was followed by two further washes in 0.2x SSC, 0.1% SDS at 65°C for 15 min each. Autoradiography was performed for one to five days at -80°C with a Kodak X-ray film and an intensifying screen.

2.8 CONSTRUCTION OF A COTTONSEED cDNA LIBRARY

2.8.1 mRNA Purification

The messenger RNA (mRNA), i.e. poly(A)⁺ RNA, was isolated from total RNA using a mRNA purification kit (Pharmacia). The key feature of this mRNA purification kit is its use of spun columns for the affinity purification of polyadenylated RNA on oligo(dT)-cellulose. This method relies on base pairing between the poly(A)⁺ residues at the 3' end of mRNAs and the oligo(dT) residues coupled to the solid support. Non-poly(A)⁺ species are washed off, and bound mRNA is subsequently eluted with a low salt buffer.

Each of the RNA samples was dissolved in 1.0 ml of TE buffer and heat denatured at 65°C for 5 minute before it was chilled on ice. 0.2 ml of sample buffer (0.5 M NaCl) was added to each of the RNA samples and mixed gently. The sample was then applied to the top of the cellulose bed in the column and it was centrifuged at 350x g for 2 min. 0.25 ml of high-salt buffer (0.5 M NaCl) was added to the column and centrifuged at 350x g for 2 min. This wash step with high-salt buffer was repeated once. The column was washed three times with 0.25 ml aliquots of low salt buffer (0.1 M NaCl) using the same centrifugation routine. To elute the bound poly(A)⁺ RNA, the column was washed four times with successive 0.25 ml aliquots of prewarmed TE buffer (65°C) and centrifuged at 350x g for 2 min after each application. The entire elute was collected in the same sterile tube. The elute was subjected to another round of purification following the above described procedures.

2.8.2 cDNA Synthesis

The synthesis of cDNA was accomplished using a cDNA synthesis kit (Pharmacia). 1-5 µg of mRNA prepared as described in above section was dissolved in 20 µl RNase-free dH₂O in an eppendorf tube and incubated at 65°C for 10 min, then chilled on ice. The heat denatured RNA, together with 1 µl DTT (1 M) solution, was added to a first-strand mix containing an oligo(dT)₁₂₋₁₈ primer and Molony Murine Leukemia Virus (MMLV) reverse transcriptase. The first strand cDNA was thus synthesised by incubation at 37°C for 1 hour.

Second-strand synthesis involved a modification of the procedure of Gubler and Hoffman (1983), in which RNaseH nicks the RNA strand of the RNA:cDNA duplex formed in the first step, and DNA polymerase I uses these nicks to replace RNA with DNA by nick translation. This process was initiated by transferring the entire first-strand reaction (33 µl) to a predisposed second-strand mix (provided by the manufacturer). It was incubated at 12°C for 1 hour and then at 22°C for 1 hour. Following second-strand synthesis, 1 µl of Klenow fragment (provided by the

manufacturer) was added and incubated at 37°C for 30 min to ensure that the ends of the cDNA are blunt. Following heat inactivation at 65°C for 10 min, equal volume (100 µl) of phenol/chloroform/isoamyl alcohol was added and mixed by vortex, then centrifuged for 3 minute. The upper aqueous layer containing the cDNA was transferred into a new tube and further purified using a spun column.

To prepare the blunt-ended cDNA for insertion into the EcoRI site of a lambda vector, an EcoRI/NotI adaptor was ligated to each end. The adaptor was composed of two oligonucleotides which form a duplex containing a phosphorylated blunt end and a nonphosphorylated EcoRI overhang, as well as an internal NotI site:

```
AATTCGCGGCCGC
      GCGCCGGCGp
```

One microlitre of EcoRI/NotI adaptors, 1 µl ATP (adenosine triphosphate) solution and 3 µl T4 DNA ligase were added to the spun column effluent (100 µl) from the previous step. Ligation of adaptors was carried out overnight at 12°C and inactivated at 65°C for 10 min. To ensure the EcoRI terminated cDNAs were phosphorylated, 10 µl of ATP solution and 1 µl of T4 polynucleotide kinase were added and incubated at 37°C for 30 min. This reaction adds the terminal phosphates required for ligation of the cDNA to vector. Following heat inactivation at 65°C for 10 min, the reaction was extracted by equal volume of phenol/chloroform/isoamyl alcohol. The upper aqueous layer (115-120 µl) was collected and applied to a second column which allowed the cDNA pass through the column, but the adaptors were retained.

2.8.3 Construction of a cDNA Library

The Lambda ZAPII system combines the high efficiency of lambda library construction and the convenience of a plasmid system with improved blue-white colour selection. Moreover, the Lambda ZAPII system allows *in vivo* rapid excision of the pBluescript® phagemid, allowing cDNA inserts to be characterised in a plasmid system. The

polylinker of the pBluescript phagemid has 21 unique cloning sites flanked by T3 and T7 promoters and 6 different primer sites for DNA sequencing.

2.8.3.1 *Ligation and packaging*

Two microlitre of cDNAs prepared as described in above section was added into an eppendorf tube together with 1 μ l of Lambda ZAPII prepared arms (1 μ g), 0.5 μ l of 10x ligation buffer (500 mM Tris-HCl pH7.5, 70 mM MgCl₂, 10 mM DTT), 0.5 μ l of 10 mM rATP (pH7.5) and 1 μ l of T4 DNA ligase (2 Weiss units). The mixture was incubated at 12-14°C overnight.

A freeze-thaw extract provided by the manufacturer was thawed between fingers until just beginning to thaw. The cDNAs were immediately added to the freeze-thaw extract. They were mixed well and incubated at room temperature (22°C) for 2 hours allowing the packaging reaction to occur. This was followed by adding 500 μ l of SM buffer (5.8 g/l NaCl, 2.0 g/l MgSO₄·7H₂O, 50.0 mM Tris-HCl pH7.5, 0.01% gelatine) and 20 μ l of chloroform, mixed gently and then centrifuged briefly to sediment the debris. The supernatant was transferred to a new tube and stored at 4°C until it was titered.

2.8.3.2 *Titering of cDNA library*

A single colony of the bacterial host XL1-Blue MRF' was transferred into 50 ml of LB medium supplemented with 0.2% maltose and 10 mM MgSO₄ in a sterile Erlenmeyer flask and grown overnight with shaking at 30°C. This lower temperature ensures that the cells will not overgrow. The cells were centrifuged in a sterile conical tube for 10 min at 2,000x rpm. The pellet was gently resuspended in 10 mM MgSO₄ to the final cell density of OD₆₀₀=0.5.

Each of a serial dilution of the packaged ligation (1, 1/10, 1/100) and 200 μ l of the host cells at OD₆₀₀ =0.5 were mixed and incubated at 37°C for 15 min. This was followed by adding 2-3 ml of cooled (48°C) top agar (0.5% NaCl, 0.2% MgSO₄·7H₂O, 0.5%

yeast extract, 1% NZ Amine, pH7.5, 0.7% agarose), 15 μ l of 0.5 M IPTG (isopropylthio- β -D-galactoside) and 50 μ l of X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside, 250 mg/ml in dimethylformamide) and plated immediately on NZY plates (0.5% NaCl, 0.2% $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.5% yeast extract, 1% NZ Amine, pH7.5, 1.5% agar). Plaques were visible after 6-8 hours in a 37°C incubator. There were about 92% recombinant plaques. The titre of the unamplified library was determined to be 1.5×10^7 plaque formation units (pfu) per millilitre.

2.8.4 Amplification of the Cottonseed cDNA Library

Twenty aliquots of the packaged mixture or library suspension containing about 5×10^4 pfu were mixed with 600 μ l of $\text{OD}_{600}=0.5$ host cells in Falcon 2059 tubes and incubated at 37°C for 15 min. Each aliquot of infected bacteria was mixed with 6.5 ml of cooled (about 48°C) melted top agar and spread evenly onto a freshly poured 150-mm plate of bottom agar. The plates were incubated at 37°C for 6-8 hours. Each of these plates was then overlaid with 8-10 ml of SM buffer and incubated at 4°C overnight with gently rocking. The bacteriophage suspension was recovered from each plate and pooled into a sterile polypropylene container and chloroform was added to a final concentration of 5%. Following a period of 15 min incubation at room temperature, cell debris was removed from the bacteriophage suspension by centrifugation at 2,000x g for 10 min. The supernatant was recovered and aliquots were stored in 7% DMSO (dimethyl sulfoxide) at -70°C. The titre of the amplified cottonseed cDNA library was checked as described above and it was determined as 2×10^9 pfu/ml.

2.9 SCREENING OF A cDNA LIBRARY

2.9.1 Plating of a cDNA Library

A cottonseed cDNA library was plated out on 20 petri dishes (150 mm) containing NZY medium to approx. 50,000 pfu/plate with 600 μ l of OD₆₀₀= 0.5 XL1-Blue MRF^r cells/plate and 6.5 ml of top agar/plate as described above. When the plaque emerged and became the right size but not joining together following approximately 8 hours of incubation at 37°C, the plates were chilled at 4°C for about 2 hours.

Hybond N⁺ membrane filter was numbered and laid, numbered side down, on the top of the plaques in each plate. The filter was then marked in three asymmetric locations by stabbing through it and into the agar beneath with a 10-gauge needle. When the filter was thoroughly wet, it was peeled from the plate and placed, numbered side up, on the top of several layers of 3MM filters saturated with denaturing buffer (1.5 M NaCl, 0.5 M NaOH) for 7 min, then neutralised in two changes of neutralising buffer (1.5 M NaCl, 0.5 M Tris-HCl, pH7.0, 0.001 M EDTA), 3 min each. The filters were washed in 5x SSC briefly and blotted dry between two layers of 3 MM filters. Filters were then fixed by laying on top of several layers of 3 MM filters saturated with 0.4 N NaOH for 10 min and rinsed in two changes of 5x SSC briefly before being blotted dry between 3MM filters.

2.9.2 Screening by DNA Hybridisation

The filters containing the replicas were soaked in 2x SSC briefly and then transferred to a hybridisation container containing enough prehybridisation buffer to cover the filters completely and incubated at 42°C for at least 4 hours with gentle shaking. The prehybridisation buffer was essentially the same with that used for DNA Southern blot analysis described in section 2.7. An α -³²P labelled DNA probe was prepared as described in section 2.6. Following prehybridisation, the α -³²P labelled DNA probe

was denatured by incubating at 95°C water bath for 5 min and chilled rapidly on ice. The DNA probe was added to the prehybridisation buffer and mixed well and allowed the hybridisation occur at 42°C for 16 hours.

When the hybridisation was completed, the hybridisation solution was removed and the filters were immediately washed in prewarmed 2x SSC, 0.1% SDS briefly at 65°C, unless otherwise stated. This was followed by two consecutive washes in 0.2x SSC, 0.1% SDS for 15 min each. Unless otherwise stated, all the washes were normally carried out at 65°C. After washing, the excess liquid was removed by blotting on Whatman 3MM paper. The filters were then placed between two sheets of plastic wrap in a cassette and exposed to X-ray film (Kodak) overnight at -80°C with an intensifying screen.

Following the development of the film, the film and filters were lined up using mark numbers and the asymmetrically located dots where needle poked through the filters. The positive plaques were identified by locating the origin of strong signals in the stock plates. A square centimetre "window" was cut from the stock plate where the putative clone lined up with the film spot using an inverted yellow tip. The plaque window was transferred into 1 ml of SM buffer and 20 µl of chloroform and mixed well by vortex. The titre was checked on a small NZY plate as described in above section so that one plate would have about 50-100 plaques. Filter lifts were taken from prechilled plates and hybridisation was carried out as described above. In the case of that the positive plaques were still too close to the background plaques, a tertiary screening was performed to obtain the isolates.

2.9.3 *In vivo* Excision of Phagemid

The ExAssist/SOLR system provided by the manufacturer is designed to allow efficient excision of the pBluescript SK(-) phagemid from the Lambda ZAPII vector. The positive plaques were cored from the stock plate and transferred to a sterile eppendorf

tube containing 500 µl of SM buffer and 20 µl of chloroform. The phage particles were released into the SM buffer by vortex. Following 2 hours incubation at room temperature, 100 µl of phage stock was taken and mixed with 200 µl of OD₆₀₀=1.0 XL1-Blue MRF' cells and 1 µl of ExAssist helper phage (>1 x 10⁶ pfu/ml) in a 50-ml conical tube and incubated at 37°C for 15 min. 3 ml of 2x YT media (1% NaCl, 1% yeast extract and 1.6% bacto-tryptone) were then added and incubated for 2 hours at 37°C with shaking. The tube was heated to 70°C for 20 min, then centrifuged for 15 min at 4,000x g. 1 µl phage stock containing the plasmid packaged as a filamentous phage particle was added to an eppendorf tube containing 200 µl of SOLR cells (OD₆₀₀ = 1.0) and incubated at 37°C for 15 min. 100 µl of the phage infected cells were plated on LB+Ampicillin plates (50 µg/ml) and incubated overnight at 37°C. All the colonies appearing on the plate should contain the pBluescript SK(-) double-stranded phagemid with the cloned DNA insert.

2.10 MAKING OF GENE CONSTRUCTS

The cDNAs encoding for Δ9 stearoyl-ACP desaturase and microsomal ω-6 desaturase cloned in the current study, namely pgh9-1 and pghD12-1 (described in Chapters Four and Five, respectively) were used to make sense and antisense gene constructs for future use in the genetic engineering of cottonseed oil.

Since the cottonseed cDNA library was prepared using a NotI/EcoRI adapter (section 2.8.2), the insert of pgh9-1 and pghD12-1 cDNAs were readily released from the pBluescript SK(-) vector by digesting approximately 1 µg of plasmid DNA with NotI at 37°C for 2 hours. The digests were then extracted with phenol/chloroform/isoamyl alcohol (25:24:1) and precipitated by adding 1/10 volume of 3 M NaAc and 2 volume of 100% ethanol. Approximately 1 µg of pGle-10 plasmid (Fig. 2.1A) containing a soybean lectin promoter and 3' lectin terminator (Cho et al., 1995) was linearised with NotI and dephosphorylated by co-incubation together with 1 unit of CIAP at 37°C for 30 min. Following heat inactivation of CIAP at 65°C for 10 min, the pGle-10 plasmid

DNA was extracted with phenol/chloroform and ethanol precipitated. A 30- μ l ligation reaction was set up in an eppendorf tube containing 1x ligation buffer (66 mM Tris-HCl pH7.6, 6.6 mM MgCl₂, 1 mM DTT and 1 mM ATP), 100 ng of the dephosphorylated linearised pGle-10 DNA and 300 ng of NotI digested pgh9-1 or pghD12-1 DNA, together with 2 unit T4 DNA ligase (Promega). Following incubation at 16°C overnight, the ligation reaction was terminated by ethanol precipitation and used to transform DH5 α competent cells as described in section 2.3. The white colonies containing pgh9-1 or pghD12-1 inserts were selected and the orientation of the cDNA relative to the lectin promoter was determined by DNA sequencing. Both sense and antisense orientations of each cDNA were selected and plasmid DNAs were prepared according to section 2.4.1 and digested with XbaI allowing the release of the whole lectin promoter-cDNA fragment. The XbaI digests were then extracted with phenol/chloroform/isoamyl alcohol (25:24:1) and ethanol precipitated. In order to convert the XbaI overhangs to blunt end termini, DNA was resuspended T4 DNA polymerase 1x buffer containing 100 μ M of each dNTP and 0.1 mg/ml acetylated BSA. Five units of T4 DNA polymerase were added and incubated at 37°C for 5 min prior to heat inactivation at 75°C for 10 min. Plasmid DNA of pBI121 (Fig 2.1B, Clontech) was digested by HindIII/EcoRI, blunt-ended and dephosphorylated as described above. A blunt-end ligation of above prepared pBI121 vector DNA and pGle-10(+cDNA) was set up in a 50- μ l reaction containing 1x T4 DNA ligation buffer (Promega) and 3 units of T4 DNA ligase (Promega) and incubated at 16°C overnight. The ligation was terminated by ethanol precipitation and electroporated into competent cell DH5 α prior to being plated out on Kanamycin-supplemented LB plates. Each combination of lectin promoter and cDNA (pgh9-1 or pghD12-1) orientation (sense or antisense) cloned into the binary vector was selected and confirmed by restriction mapping and DNA sequence determination.

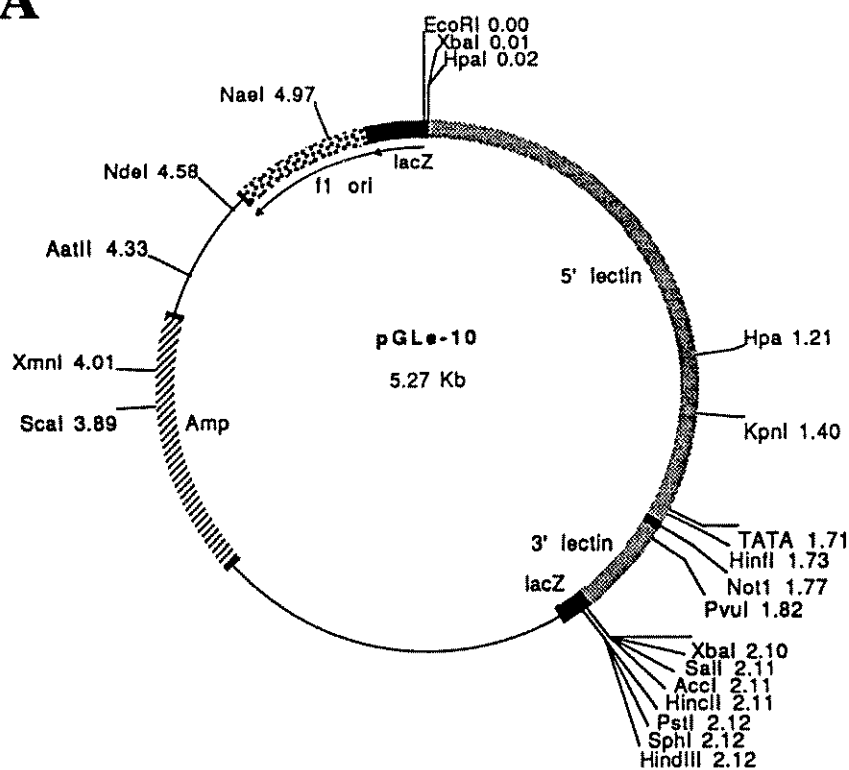
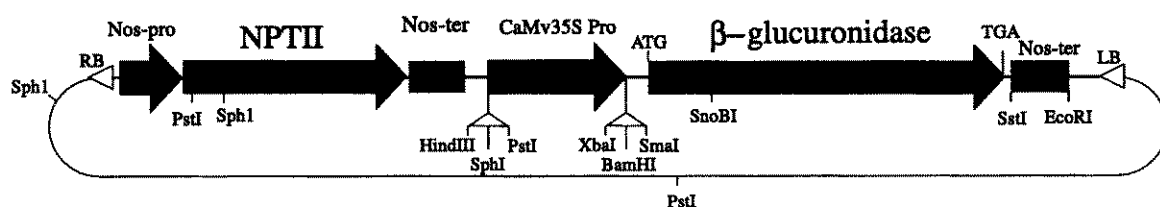
A**B**

Fig. 2.1 Plasmids used to make the sense and antisense gene constructs of cotton fatty acid desaturase gene driven by a lectin promoter.

(A). pGLE-10 (Cho et al., 1995)

(B). pBI121 (Clontech)

Chapter Three

A Preliminary Study on the Oil Content and Fatty Acid Composition in Developing and Mature Cottonseeds

3.1 INTRODUCTION

Previous studies have found considerable natural variability for oil content and fatty acid composition in seed oil among some wild relatives of cotton (Carter et al., 1966; El-Nockrashy et al., 1969; Khattab et al., 1977). There are also a few reports on the analyses of commercial varieties, mostly on samples from the United States cottonbelt (Cherry, 1983). It has been observed that cultivar, location, year, and their interactions are highly significant sources of variation associated with cottonseed oil content and its composition (Pope and Ware, 1945; Bailey, 1948; Stansbury et al., 1954, 1956; Cherry et al., 1970; Pandey and Thejappa, 1975; Lawhon, et al., 1976; Cherry et al., 1978, 1979, 1981; Cherry and Berardi, 1983; Cherry, 1983; Kohel and Cherry, 1983). Although such investigations allow certain trends to be observed, the data are difficult to interpret satisfactorily because they are limited in geographic scope and represent the composite effects of varietal and ecological factors. The desirable mutant types discovered in other oilseeds, such as the high stearic soybean (Graef et al., 1985), high palmitic sunflower (Ivanov et al., 1988) and high oleic rapeseed (Auld et al., 1992),

have not been identified or produced in cottonseeds. Despite many years of interest in improving cottonseed quality, research on the quality of cottonseed oil had little impact on the improvement of cotton cultivars.

The cotton industry in Australia is a relatively young industry, the genetic variation in oil content and fatty acid composition of cottonseed oil of elite Australian cultivars has not been systematically studied. Owing to the competition pressures from other oilseeds and the increasing production of cottonseeds as a byproduct of cotton fibre, there is increasing awareness and demand to modify the cottonseed constituents, especially fatty acid composition in seed oil towards a value-added commodity.

The present study was initiated as a preliminary survey of the variation in oil content and fatty acid composition of a collection of several current Australian cultivars, a few other representative genotypes and wild *Gossypium* species. The other objective in this study is to describe the accumulation of oil and its fatty acid composition in developing cottonseed. The experiment was rather simply designed and the information revealed in this study may only be used as a coarse reference for future investigations of Australian cultivars concerned with genetic modification of oil accumulation and fatty acid composition in cottonseed oil.

3.2 MATERIALS AND METHODS

3.2.1 Plant Materials

The seeds of four Australian native *Gossypium* species, including *G. australe*, *G. bickii*, *G. robinsonii* and *G. sturtianum* were purchased from Nindethana Seed Service, WA. Seeds of the remaining sixteen entries were kindly provided by Dr. Greg Constable of Australian Cotton Research Institute at Narrabri, NSW.

Several plants of *G. hirsutum* cv Siokra were grown in glasshouse in Plant Breeding Institute, Cobbitty, Sydney University. The growth temperature was with 28/15°C

day/night and natural day light was extended with fluorescent light up to 16 hours photoperiod per day. Throughout flowering, flowers were tagged at anthesis to use as indicators for the chronological age of the developing cotton embryos. Embryos were harvested at 5 day intervals until the seeds were physiologically mature.

3.2.2 Lipid Extraction

The total lipids in cottonseed or seed oil were extracted by a modification of the method of Folch et al. (1957). Dry cottonseeds were ground in a coffee grinder. Immature embryos were harvested immediately before grinding into powder in liquid nitrogen with pestles and mortars. Approximately 4 grams ground powder from mature seed samples were weighed out and the exact weights were recorded. Because of the limited availability of immature embryos, only 1-2 gram samples were used. They were placed into an Erlenmeyer flask. Total lipids were then extracted with 60 ml of Folch reagent (2:1 v/v chloroform:methanol solution, 15 ml per gram of sample). The extracts were filtered on Whatman No. 1 filter paper into a 100-ml graduated cylinder and made up to 60 ml. Then they were washed with 12 ml of saline solution (0.88% NaCl, 20% volume of the extract) and the phases were allowed to separate. The upper phase, containing water, methanol and water soluble material was siphoned off and discarded. The lower phase of chloroform and lipids was further washed with a chloroform:methanol:saline (3:47:48 v/v) solution. Again phases were allowed to separate. The final volume of the lower chloroform:lipid was recorded and the upper layer siphoned off and discarded. 10 ml of the chloroform extract and 1 ml of internal standard, heptadecanoic acid (C17:0, 5.196 mg/ml) were transferred into a preweighed culture tube with a screw cap. The sample was evaporated to dryness under a nitrogen gas stream on a 70°C heating block. The amount of crude lipids per gram of sample was determined by weighing the tube containing the dried sample.

Five millilitres of the methylating reagent, a mixture of boron trifluoride:methanol:hexane (35:45:20) was added to the culture tube containing dried lipid and capped

tightly with a teflon lined crew cap. Then it was incubated in a 90°C water bath for 45 min with frequent shaking. After cooling down to room temperature, 5 ml of water and 5 ml of hexane were added to the mixture. Following shaking and phase separation, the top hexane layer containing the fatty acid methyl esters was transferred into a new tube and used in the GLC analysis.

3.2.3 Fatty Acid Analysis by Gas-Liquid Chromatography (GLC)

The fatty acid composition of the lipids was determined using a Varian Gas Liquid Chromatography, model 3400, equipped with a fused capillary column DB624 (J & W Scientific). The following program was used in the analysis: the column temperature was held at 70°C for 3 min, then increased to 160°C at a rate of 30°C/min and held for 10 min. It was then increased to 230°C by 7°C/min and held for 4 min. The injector and detector temperature were 220°C and 240°C, respectively. The nitrogen flow rate was 30 ml/minute.

A fatty acid standard containing the fatty acids lauric acid (12:0), myristic (14:0), palmitic (16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2), α -linolenic (C18:3), arachidic (C20:0), behenic (C22:0), eicosadienoic (C20:2) and lignoceric (C24:0) was used. Fatty acid peaks were identified by comparing the fatty acid methyl ester peaks and retention times of standards with sample peaks. An electronic integrator was used to calculate the total area of peaks and the area of each fatty acid peak was expressed as a percentage of the total area.

3.2.4 Statistical Analysis

The oil contents and fatty acid compositions of dry cottonseeds were determined from four independently prepared seed samples per entry. Three replicates of immature embryo samples were prepared and used for fatty acid analysis. Analysis of variance (ANOVA) was performed using SuperANOVA software (v1.11, Abacus concepts, Inc.

1989-1991). The correlation analysis was carried out using StatView 4.0 software (Abacus concepts, Inc. 1992).

3.3 RESULTS AND DISCUSSION

3.3.1 Oil Contents in the Major Elite Australian Cotton Cultivars and Some Other Genotypes and Species

Fig. 3.1 illustrates a typical chromatography pattern of fatty acids in cottonseed oil detected by GLC in this study. Palmitic acid (C16:0) represents the major saturated fatty acid, while low level of other saturated fatty acids including myristic acid (C14:0) and stearic acid (C18:0) were also detected. Oleic acid and linoleic acid were the two major unsaturated fatty acids; while a very low level of palmitoleic acid (C16:1) was also detectable. Stearic acid is desirable for the production of margarine and cocoa butter-like fat. The low level of stearic acid in cottonseed oil is probably because of the strong activity of a Δ^9 stearyl-ACP desaturase which converts most of the stearic acid to oleic acid which is further desaturated to form linoleic acid by a microsomal ω -6 desaturase. Linoleic acid accounts for more than 50% of total fatty acid composition, indicating the strong activity of the microsomal ω -6 desaturase in cottonseeds. Perhaps the most desirable feature of this oil is its extremely low linolenic acid content which is unsuitable for food applications due to its instability and deterioration of flavour associated with autoxidation (Smouse, 1979; Green, 1986). The minor-peaks detected following linoleic acid in Fig. 3.1 represent α -linolenic (C18:3), arachidic (C20:0), gadoleic (C20:1) and lignoceric (C24:0) acids, respectively.

It may be also noted that the current GLC analyses did not differentiate cyclopropenoid fatty acids (CPFA) from other unsaturated fatty acids. Previous studies have demonstrated that cottonseed oil contains less than 1% cyclopropenoid fatty acids (CPFA), mainly sterculic and malvalic acids, which contain one double bond at the site of a propene ring, either at the 9-10 or 8-9 position (Phelps et al., 1964; Allan et al.,

1967). Poultry and other animal feeding studies have shown CPFA to change the ratio between stearic and oleic acids; as a result, CPFA cause the hardening of fats in egg yolk and milk (Johnson et al., 1967). CPFA is unstable during the standard lipid extraction process, but derivatives of CPFA have been quantitated by modified GLC technique and high performance liquid chromatography (Fisher and Schuller, 1981; Fisher and Cherry, 1983; Pandey and Subrahmanyam, 1988).

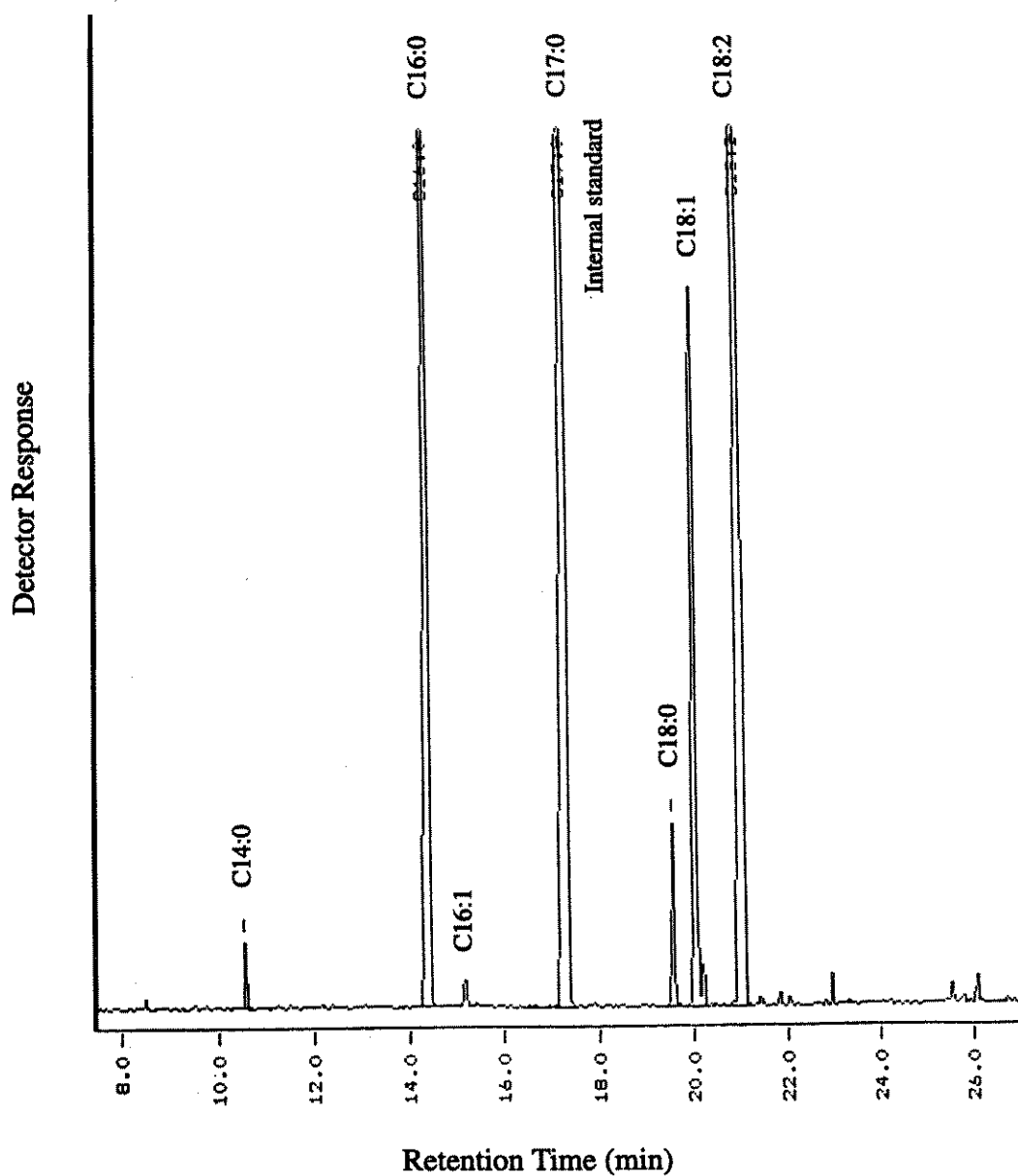


Fig. 3.1 A typical chromatography chart of the fatty acid analysis of cottonseed oil in the current study. A C17:0 fatty acid was used as internal standard.

The mean values for percentage oil content and the component fatty acids in the cottonseed oils from 20 entries are reported in Table 3.1. Student-Newman-Keuls test for the oil content and component fatty acids of various entries indicated that species or genotypes have a significant influence on the oil content and the majority fatty acid compositions of cottonseed oil (except for C14:0) at the 5 percent level of probability.

The four Australian wild diploid *Gossypium* species contained significantly lower oil content than the remaining entries examined. They were also found to contain significantly higher level of linoleic acid and lower level of palmitic and oleic acids. In contrast, the only A-genome diploid species, *G. arboreum*, was found to contain significantly higher level of oleic acid, but significantly lower level of linoleic acid. This result is largely in agreement with previous studies on the chemical survey in genus *Gossypium* (El-Nockrashy et al., 1969; Khattab et al., 1977).

However, I have been aware of that not all the entries were grown under the same environment and therefore the environmental effects on the entries could not be evaluated. It appears to be a general phenomenon that the oil content and fatty acid composition can be greatly influenced by growth temperature (Neidleman, 1987). Undoubtedly, some of the variability observed here was caused by differences in environments at different locations. Therefore, the current data and analysis only allow a certain trend to be observed, but further studies are needed to be carried out under more stringent conditions.

The major Australian elite cotton cultivar SicalaV2, SiokraL22, SiokraL23 and CS50 had the highest oil content, whereas Sicala33 and Sicala34 were found to contain the moderate oil content, no significant difference from the American cultivar, Acala R. As far as the fatty acid composition is concerned, SiokraL22, SiokraL23, Sicala33 and Sicala34 contain relatively high levels of palmitic acid and moderate levels of unsaturated fatty acids.

Table 3.1 Oil Content and Fatty Acid Analysis in Seed Oil of Some Cotton Cultivars and a Few Other *Gossypium* Species

Entries	Lipid Content (%)	Fatty Acid (%)					
		14:0	16:0	16:1	18:0	18:1	18:2
<i>G. australe</i>	16.2b	0.7a	21.5a	1.6b	1.7a	11.0b	62.1f
<i>G. bickii</i>	14.6a	0.8a	20.6a	0.9a	3.5c	8.6a	64.1g
<i>G. robinsonii</i>	17.5b	0.9a	22.8abc	0.6a	2.7b	10.4ab	60.0e
<i>G. sturtianum</i>	17.3b	1.1a	20.4a	0.5a	2.6b	11.2b	62.5fg
<i>G. arboreum</i>	22.4de	0.7a	21.0a	1.0a	3.6c	27.2e	45.3a
<i>G. barbadense</i>	21.1cd	0.7a	24.4bcd	0.7a	2.2ab	19.4d	52.0bc
Acala R	23.0ef	0.6a	24.4bcd	0.5a	2.3ab	16.9cd	54.4cd
BAR7/8	21.1cd	0.9a	23.9bc	0.6a	2.3ab	15.6c	56.1d
DP90	22.0cde	0.7a	26.9d	0.6a	2.3ab	16.3cd	52.5bc
MCU-5	20.6c	0.7a	22.4ab	0.8a	2.1ab	16.4cd	56.1d
lintless 28	23.7efg	0.8a	24.3bc	0.5a	2.6b	18.7cd	52.6bc
lintless 53	23.7efg	0.8a	25.5cd	0.5a	3.7c	18.3cd	50.6b
Sicala33	23.2ef	0.8a	24.6bcd	0.7a	2.7b	16.7cd	53.4c
Sicala34	23.0ef	0.7a	24.6bcd	0.5a	2.5b	16.7cd	54.0cd
CS50	25.0gh	0.6a	24.1bc	0.4a	2.5ab	17.7cd	54.1cd
CS65	23.8efg	0.9a	22.6ab	0.6a	2.5ab	16.3cd	56.4d
Siokra L22	24.5fgh	0.8a	24.9bcd	0.7a	2.7b	17.2cd	53.1bc
Siokra L23	24.7fgh	0.7a	25.1bcd	0.7a	2.5ab	17.5cd	52.7bc
Siokra 1-4	25.5h	0.7a	24.4bcd	0.5a	2.4ab	17.5cd	54.2cd
Sicala V-2	26.0h	0.7a	23.8bc	0.5a	2.4ab	17.9cd	54.1cd

*The least significant differences are calculated for the 5% level of probability. Within columns means having a letter in common are not significantly different.

The co-relationships among various fatty acids in cottonseed oil are summarised in Table 3.2. No significant correlations were found among saturated fatty acids. In contrast, the correlation coefficient between oleic and linoleic acids was estimated to be -0.922, which is very highly significant and reflects the precursor-product relationship. The highly negative correlation between oleic and linoleic acid was also consistent with reports from other oilseed species (Fernández-Martínez et al., 1993; Seiler, 1994; Singh et al., 1998).

The levels of two monounsaturated fatty acids, palmitoleic and oleic acids were also found to be highly negatively correlated. This may reflect the competition between substrates in the biochemistry process of fatty acid desaturation to form the monounsaturated fatty acids. Although the $\Delta 9$ stearoyl-ACP desaturase has a relatively high specificity for the stearic acid substrate, to a less extent it also catalyse the desaturation of palmitic acid. The major saturated fatty acid, palmitic acid was also found to have moderate correlations with palmitoleic acid and linoleic acid, respectively.

Table 3.2 Correlation Coefficients of Fatty Acid Composition of Cottonseed Oil

	myristic acid	palmitic acid	palmitoleic acid	stearic acid	oleic acid
palmitic acid	-0.172	-	-	-	-
palmitoleic acid	0.156	-0.367	-	-	-
stearic acid	0.156	-0.161	-0.107	-	-
oleic acid	-0.161	0.172	-0.85	0.022	-
linoleic acid	0.115	-0.481	0.157	-0.110	-0.922

3.3.2 The Changes of Fatty Acid Composition in Developing Cottonseeds

The oil accumulation and changes of fatty acid composition in developing cottonseed embryos aged 15-65 days after anthesis (DAA) is illustrated in Fig. 3.2. There was very little oil being synthesised before the sampling time (15 DAA). In the beginning of sampling, i.e. between 15-20 DAA, oil synthesis was also relatively slow. There was a relatively rapid increase of oil content between the 20-35 DAA. The percentage values were essentially constant during the remainder of the seed development (35-65 DAA). Such an observation is consistent with previous reports (El-Nockrashy et al., 1975; Kajimoto et al., 1979).

The six species of fatty acids, myristic (14:0), palmitic (16:0), palmitoleic (16:1), stearic (18:0), oleic (18:1) and linoleic (18:2) acids detected in developing embryos are generally recognised as the fatty acid constituents of cottonseed. The rate of synthesis of linoleic acid was greater than those in other fatty acids. A rapid increase in synthesis of linoleic acid was observed between 15 to 30 DAA from 22% at 15 DAA to almost 50% at 30 DAA, concomitant with the rapid increase of percent oil content. There was only a slight increase thereafter. There was continuous decrease in the percentage of palmitic acid between 15-35 DAA and the decrease trend became less obvious in late development stage. Interestingly, oleic acid did not increase as did the linoleic acid. Instead, oleic acid decreased gradually throughout development of the cottonseeds, but without a rapid fall. Stearic acid decreased from about 9% at 15 DAA decrease to 2.5% at 30 DAA and maintained that level thereafter. The other two minor fatty acid species, i.e. myristic and palmitoleic acids remained at relatively low level throughout the period of embryo development.

From the data presented here it appears that the percentage values of oil content and fatty acid composition are determined at a fairly early stage of development (20-35 DAA). Since environmental conditions play a key role in oil synthesis and fatty acid composition, the temperature and planting practices during the early stage of

development undoubtedly influence the fatty acid composition of cottonseed oil and also the final oil percentage. During the last 30 days of embryo development, when percentage oil and fatty acids maintain their relative proportion, these components continue to increase in absolute amounts as dry weight continues to increase. In fact, the majority of the total oil found in the mature cottonseed was synthesised during this period of time. Therefore, this is a very important and critical time in cottonseed production. The constant percentages of oil content and fatty acid composition during embryo development also indicate that oil synthesis and fatty acid synthesis is rigidly controlled in cottonseeds.

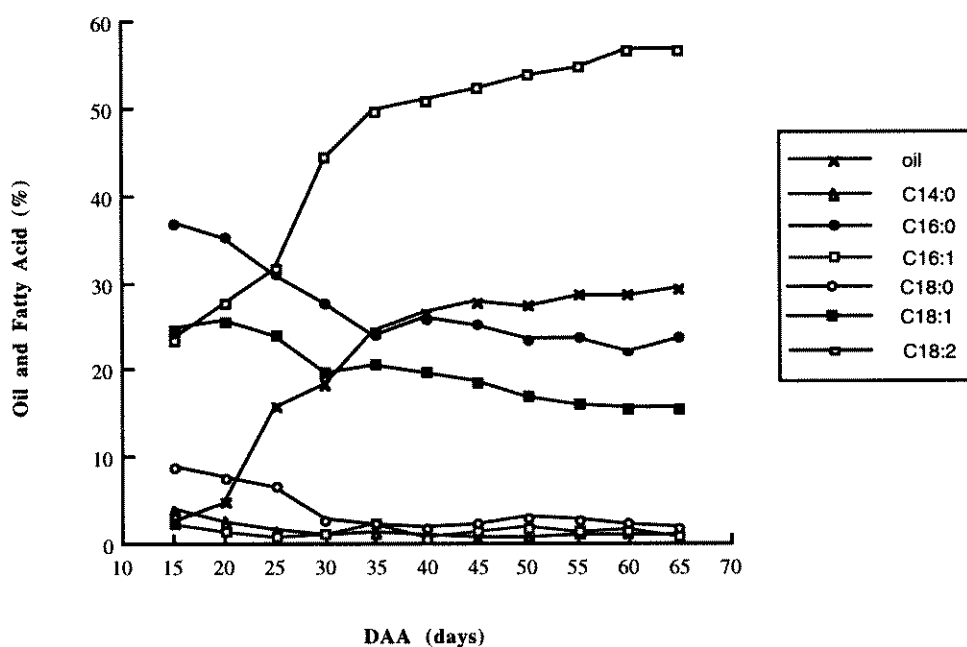


Fig. 3.2 Seed oil accumulation and changes of fatty acid composition in developing cottonseed embryos. Samples of cottonseed embryos were collected and analysed between 15 to 65 DAA at 5 day intervals.

3.4 SUMMARY

The present study indicates that there are some, but limited variations of oil content and fatty acid composition among the elite Australian cotton varieties. Although not conclusive, it appears that the wild Australian native *Gossypium* diploids have rather different fatty acid profiles. Utilisation of wild *Gossypium* germplasm coupled with embryo rescue, careful selection of cultivar and improved planting conditions may improve cottonseed oil yield and quality. However, in order to achieve dramatic alteration of fatty acid composition without affecting fibre quality so that cottonseed oil can continue to compete in the world vegetable market, genetic engineering may prove to be more effective. The current study also illustrates the pattern of oil accumulation and changes of fatty acid composition in developing cottonseed embryos. This information was taken into consideration in the subsequent experiments of RNA isolation from immature embryos, cDNA library construction and gene cloning.

Chapter Four

Isolation and Characterisation of a $\Delta 9$ Stearoyl-ACP Desaturase cDNA and Its Genomic Structural Gene from *G. hirsutum*

4.1 INTRODUCTION

In plants, fatty acid biosynthesis occurs in the chloroplasts of green tissue or in the plastids of non-photosynthetic tissues (Harwood, 1988; Ohlrogge and Browse, 1995). The primary products in most plants are ACP esters of the saturated palmitic and stearic acids. $\Delta 9$ Stearoyl-ACP desaturase (EC 1.14.99.6) is a plastid-localised soluble desaturase that introduces the first double bond into stearic acids, and palmitic acid to a lesser extent, resulting in the corresponding monounsaturated fatty acids (Nagai and Bloch, 1966). Most plants, including cotton, lack other desaturases that act on the 18:0 level. Therefore, the ratio of saturated to unsaturated fatty acids in higher plants is controlled mainly by the $\Delta 9$ stearoyl-ACP desaturase activity, although there are some related acyl-ACP desaturases specific for various chain length and double bond positions (Cahoon et al., 1992; 1994; 1997a; Cahoon and Ohlrogge, 1994; Schultz et al., 1996).

The $\Delta 9$ stearoyl-ACP desaturase normally has high activity, relative to KASII which catalyses the formation of stearoyl-ACP by extending palmitoyl-ACP, so that stearate

seldom accumulates in plants. Plant tissues such as the cocoa pods from *Theobroma cacao* are an exception which, of course, leads to the high price of cocoa butter. It has been demonstrated that antisense downregulation of the activity of $\Delta 9$ stearoyl-ACP desaturase in *B. napus* results in high stearate canola oil (Knutzon et al., 1992b).

The availability of a cloned endogenous $\Delta 9$ stearoyl-ACP desaturase gene is a prerequisite for genetic engineering of cottonseed oil towards the high stearic vegetable oil for the production of margarine and high value cocoa butter substitute. In this chapter, the isolation and characterisation of a cDNA and its corresponding structural gene encoding the $\Delta 9$ stearoyl-ACP desaturase of cotton is described.

4.2 MATERIALS AND METHODS

4.2.1 Oligonucleotide Design and Reverse Transcriptase-PCR (RT-PCR)

A pair of degenerate PCR primers were synthesised that contained some possible codons corresponding to the open reading frames (ORFs) of peptide sequences among $\Delta 9$ stearoyl-ACP desaturase cDNAs from castor bean (Shanklin and Somerville, 1991), cucumber (Shanklin et al., 1991) and flax (Singh et al., 1994). The sequences of the primers are as follows (degenerate positions are in parentheses): D9S, 5'- ATGGC-(G/T)CT(C/G)A(A/G)GCT(C/G/T)-CAT(C/G)C -3'; D9A, 5'- TCA(G/C)AG(C/T)TT-(C/A)AC(T/C)TG(T/C)CTAT-3'.

The isolation of total RNA from cottonseed embryos 25-35 DAA and subsequent mRNA purification were carried out according to the procedures described in sections 2.4.5 and 2.8.1, respectively. Single stranded cDNA template for reverse transcriptase-PCR (RT-PCR) was synthesised at 45°C from poly(A)⁺ RNA using a Superscript kit (Gibco BRL) according to the manufacturer's instruction. PCR was carried out in a 50- μ l reaction containing 50 pmol of each primer and the single-stranded cDNA derived from 20 ng of mRNA. The mixture was heated to 96°C for 5 min prior to PCR reaction. Three units of the AmpliTaq[®] DNA polymerase (Perkin-Elmer Cetus) was then added.

PCR was performed in a Corbett FTS960 thermal cycler using the 40 cycles of the following program: 94°C for 30 sec, 48°C for 30 sec and 72°C 1 min and 10 sec. This was followed by a 10 min extension at 72°C. PCR product was purified using a Wizard PCR Prep Kit (Promega) and cloned into a T-vector (Promega) according to the manufacturer's instructions. The identity of the cloned PCR product as a partial cDNA of Δ^9 stearoyl-ACP desaturase was confirmed by DNA sequencing. The partial cDNA insert was then excised from the purified plasmid DNA and radio-labelled with [α -³²P]dCTP as described in section 2.6.

4.2.2 Isolation of a cDNA from a Cottonseed cDNA Library

A cDNA library was prepared using the λ ZAPII vector (Stratagene) and poly(A)⁺ RNA isolated from the embryo tissues of developing seed of *G. hirsutum* cv Deltapine-16 as described in section 2.8 in Chapter Two. Approximately 5×10^5 Pfu were screened with the above described [α -³²P]dCTP-labelled PCR fragment as a DNA probe. Hybridisation and following washes were carried out according to section 2.9.2. Positive plaques selected in the primary screening were purified by two more screenings. Phagemid containing the Δ^9 stearoyl-ACP desaturase cDNA was excised *in vivo* using ExAssist/SOLR system as described in section 2.9.3. The DNA sequence was determined using an ABI 373 DNA sequencer.

4.2.3 Isolation of the Structural Gene from a Cotton Genomic DNA Library

A genomic DNA library constructed in the λ GEM (Promega) system with Sau3A partial digested genomic DNA from *G. hirsutum* cv Deltapine-16 was a gift from Drs Bruce Lyon and Danny Llewellyn. Approximately 1×10^5 pfu generated in *E. coli* strain K803 were screened using an [α - ^{32} P]dCTP-labelled DNA fragment corresponding to the 3' untranslated region (UTR) of the above obtained cDNA. Plaque screening was performed using Hybond N⁺ (Amersham) using a protocol essentially the same with the cDNA library screening procedure described in sections 2.9.1 and 2.9.2.

The lambda DNA was isolated and purified according to the following procedures. A single colony of K803 was grown in 10 ml LB+10 mM MgSO₄ and 0.2% maltose overnight at 37°C. 50 μ l of K803 culture were added to 2.5 ml of top agar at 48°C, mixed well and laid on the top of the bottom agar in a 90 mm Petri dish. 1-6 drops of the purified lambda phage were added onto the solidified top agar and the plate was incubated overnight at 37°C. When the plaques appear, they were taken with an inverted yellow tip and transferred to 500 μ l SM buffer containing a few drops of CHCl₃ and incubated at 37°C water bath for 2.5 hours.

Fifty microlitre of the above lambda phage were added to a 500 μ l of overnight culture of K803 cells and incubated at 37°C for 15 min before 20 ml LB medium supplemented with 10 mM MgSO₄ were added. Following growing at 37°C for at least 3 hours until the culture became clear, four drops of chloroform and 1.17 g NaCl were added to the cleared culture and incubated at 37°C in the shaker for 30 min. Cell debris was pelleted by centrifugation at 18,000 rpm (Sorvall ss34) for 30 min at 4°C. The supernatant was transferred into a new tube containing 5 ml of 50% PEG₈₀₀₀, mixed well and left on ice for about 60 min prior to centrifugation at 10,000x rpm for 20 min. The pellet was resuspended in 750 ml SM buffer and extracted with equal volume of chloroform and the aqueous phase was transferred into a discontinuous gradient of CsCl.

The discontinuous gradient was comprised of 0.5 ml of 5 M CsCl (bottom) and 1 ml 3 M CsCl (top). Centrifugation was carried out at room temperature for 2 hours at 35,000 rpm (TL100, swing-out rotor). Approximately 300 µl of the interface was collected with a syringe into a 15 ml Corex tube. 30 µl of 2M Tris-HCl (pH8.5)/0.2 M EDTA, 300 µl formamide and 300 µl H₂O were added and mixed well, then the lambda DNA was precipitated by adding 1.8 ml of ethanol and incubated at -20°C overnight. Following a centrifugation at 10,000 rpm using a ss34 rotor for 15 min at 4°C. The pellet was resuspended in 175 µl TE and 4 µl of 20% SDS and 20 µl freshly prepared proteinase K (1 mg/ml) were added and incubated at 55°C for 1 hour. Following two consecutive extractions with phenol/chloroform/isoamyl alcohol (25:24:1), DNA was precipitated with ethanol and rinsed with 70% ethanol. After drying briefly in a speedyvac, the lambda DNA was resuspended in 20 µl TE.

The clones were mapped by restriction digestion with EcoRI, BglII, BamHI and XhoI, followed by Southern blot hybridisation under the conditions described in section 2.7. Subsequently, EcoRI-digested DNA fragments were subcloned into the pBluescriptIIKS⁻ vector (Stratagene) and a subclone harbouring an approximately 6 kb insert containing the entire gene was fully sequenced from both directions.

4.2.4 RNA Northern Blot Analysis

Total RNA was extracted from young leaves and immature embryos of 25, 30, 36, 45 days after anthesis (DAA) of *G. hirsutum* cv Deltapine-16 according to the protocol described in section 2.4.5. The RNAs were denatured with formaldehyde, resolved by electrophoresis in 1% agarose gel containing formaldehyde, transferred to Hybond-N⁺ nylon membranes according to section 2.5.3 and probed with an [α -³²P]dCTP-labelled 3' UTR fragment of the *G. hirsutum* $\Delta 9$ stearyl-ACP desaturase cDNA. The hybridisation was carried out under standard condition described in the section 2.7 in Chapter Two.

4.2.5 Genomic DNA Southern Blot Analysis

Genomic DNA extraction from young leaves of *G. barbadense*, *G. hirsutum* (cvs Deltapine-16 and Siokra), *G. herbaceum* and *G. raimondii* and further purification by CsCl gradient procedure were carried out as described in sections 2.4.3 and 2.4.4, respectively. Ten microgram DNA of each sample was digested with EcoRI, HindIII or XbaI and electrophoresed through a 0.7% agarose gel. The gel was blotted onto Hybond-N⁺ nylon membrane (Amersham). The filters were probed with [α -³²P]dCTP-labelled DNA fragments corresponding to the entire coding region or a 3' UTR fragment. Hybridisation and washes were carried out under the standard conditions as described in section 2.7.

4.3 RESULTS

4.3.1 Cloning and DNA Sequence Analysis of a cDNA pgh9-1

Initially, an approximately 1 kb DNA fragment consisting of the coding region of a $\Delta 9$ stearyl-ACP desaturase gene was obtained *via* RT-PCR. This PCR product was subsequently cloned and its identity was confirmed by partial DNA sequence determination. This PCR fragment was then used as a DNA probe to screen a cottonseed cDNA library under high stringency conditions. Ten plaques which showed strong positive signals in the primary screening were selected and further purified by secondary and tertiary screenings prior to phagemid excisions and DNA sequencing. Since all of these clones turned out to be identical, based on the partial sequencing and restriction mapping. Only one clone, designated as pgh9-1, was fully sequenced. The nucleotide sequence of pgh9-1 has been deposited to EMBL GenBank under the accession number X95988.

CGAAAAGAAAAATGGCTTTGAATTTTAATGCCATCGCCTCGAAATCTCAGAAGCTCCCT	60
M A L N F N A I A S K S Q K L P	
TGCTTTGCTCTTCCACCAAAGGCCACCCCTTAGATCTCCCAAGTTTTCATGATCTCCACC	120
C F A L P P K A T L R S P K F S M I S T	
ATTCTTCTGGCTCCAAAGAGGTTGGGAATCTGAAAAAGCCTTTCACGCCTCCAAAGAG	180
I P S G S K E V G N L K K P F T P P K E	
GTGCCTGTTTCAGATCACCCACTCCATGCCGCTCACAAGATTGAGATCTTTAAATCTTTG	240
V P V Q I T H S M P P H K I E I F K S L	
GAGGGCTGGGCTGAGAACAACATTCTGACTCACCTCAAACCAGTTGAGAAATGTTGGCAA	300
E G W A E N N I L T H L K P V E K C W Q	
CCCGCGACTTTCTTCCAGATCCTAATTCTGATGGATTTCATGAGCAAGTCAAAGAGCTT	360
P A D F L P D P N S D G F H E Q V K E L	
AGGGAAAGGGCAAAGGAGATCCCAGATGATTACTTTGTAGTTTGGTTGGTGATATGATC	420
R E R A K E I P D D Y F V V L V G D M I	
ACCGAGGAAGCCCTTTCAACTTATCAAACAATGCTTAATACCTTGGATGGAACCTCGTGAT	480
T E E A L S T Y Q T M L N T L D G T R D	
GAGACAGGTGCTAGCCTTACCCTTGGGCCATTGGACCAGGGCTTGGACTGCTGAAGAA	540
E A T G A S L T P W A I W T R A W T A E E	
AACGGCATGGTGATCTGCTTAATAAGTATCTCTACTTGTCTGGGAGAGTGAGATGAGG	600
N R H G D L L N K Y L Y L S G R V D M R	
CAAATTGAGAGGACAATCCAGTACTTGATTGGATCGGGAATGGATCCTCATACAGAGAAT	660
Q I E R T I Q Y L I G S G M D P H T E N	
AGTCCTTACCGAGGATTTCATATATCTTCGTTCCAAGAAAGGGCAACTTTTATTTCCCAT	720
S P Y R G G F I Y T S F Q E R A T F I S H	
GGGAATACAGGCGCTGGCTATAGGAGTATGGGATATTAACCTTGGCTCAAATTGTGGT	780
G N T G R L A K E Y G D I N L A Q I C G	
AGCATTCCTCAGATGAGAAGCGCCACGAGACAGCCTATACCAAATCGTTGAAAAGCTG	840
S I A S D E K R H E T A Y T K I V E K L	
TTTGAGATTGATCTGATGAAACAGTCTCGGCATTGCTGACATGATGAAGAAGAAAATC	900
F E I D P D E T V L A F A D M M K K K I	
GCCATGCCGGCTGAGTTTCATCTATGATGGCAGAGATTATAACTTATTTGACCACTACTCA	960
A M P A E F I Y D G R D Y N L F D H Y S	
GCTGTTGCCCAAAGAATCGGGGTTTACACTGCTAAGGACTATGTTGATATAGTAGAGCAC	1020
<u>A V A O R I G V Y T A K D Y V D I V E H</u>	
CTGGTGGATCGATGGAAGGTGAAGGAGCTAGCTGGGCTTTTCAGCCGAGGGGCGTAAAGCT	1080
L V D R W K V K E L A G L S A E G R K A	
CAGGACTACTTGTGTTTCCTTCGAGAATTAGAAGGTTAGAGGAGAGAGCCCAAGAA	1140
Q D Y L C S L P S R I R R L E E R A Q E	
AAGGCCAAGGAAGCACCCAGTGTCCCATTCAGTTGGATATTTGATAGAGAAGTGAACTT	1200
K A K E A P S V P F S W I F D R E V K L	
TAGGTCATGAAATACAGTTAAGACTCCTGCAATGCATTTGAGGAAACAAACACGAAGAAG	1260
*	
CTGAATGCCAACTTCTCTTTATATATCCGATGTAATAGAGGTTGTATATGTAACAGGAGG	1320
AATTGCGTGGCTTTGGTTAGGGTAGCACATGTTTCTGGATGTGTTGTGCTTTAAAAA	1380
TAATGCCGATAGCGGCAGCTGTGATAGTTTATAGATGTTTGTGTTTCATAATGTCTGTTATA	1440
TCGTTGTACGAGTAGTATGTGTTGTTTGTGTTGAAACAATCTTCATATCTTAGTGATAAA	1500
TGATAATGCTGTGTAGTCATAGTTTGTAGTTTGCATAAAAAAAAAAAAAA	1553

Fig. 4.1 Nucleotide and deduced amino acid sequences of a cDNA, designated as pgh9-1, encoding the $\Delta 9$ stearoyl-ACP desaturase of *G. hirsutum*. The translation start site ATG is underlined. An asterisk indicates the stop codon. The two iron-binding motifs separated by about 100 amino acids are in bold. The likely cleavage site for transit peptide is indicated by an arrow. The motif representing the fatty acid desaturase family II signature is also underlined.

Pgh9-1 is 1553-bp in length, including 5' (12-bp), 3' (350-bp) UTRs and a 1191-bp ORF which encodes a polypeptide of 397 amino acid residues. The sequence surrounding the putative ATG start site (AAATGGCT) has 78% identity to the consensus for plant translation initiation proposed by Lütcke et al. (1987). The identity of pgh9-1 as a $\Delta 9$ stearyl-ACP desaturase cDNA was confirmed by its sequence homology with other $\Delta 9$ stearyl-ACP desaturase cDNAs in GenBank database.

The deduced *G. hirsutum* $\Delta 9$ stearyl-ACP desaturase contains an apparent plastid transit peptide sequence at its N-terminus, as is present in all previously characterised acyl-ACP desaturases. This is consistent with its nature as a plastidial enzyme encoded by a nuclear gene. Based on homology with castor bean (Knutzon et al., 1991), it is predicted that amino acid 32 of the polypeptide deduced from pgh9-1 is the most likely cleavage site for the plastid transit peptide (indicated by an arrow in Fig. 4.1). Also, residues immediately preceding amino acid 33 conform closely to those of the consensus cleavage site for plastid transit peptides proposed by Gavel and von Heijne (1990). Therefore, it is likely that the pgh9-1 ORF codes for a precursor protein with a 32 amino acid transit peptide to translocate this enzyme into the proplastid and subsequent cleavage of the transit peptide produces a mature protein of 365 amino acid residues. Similar to previously characterised plant $\Delta 9$ stearyl-ACP desaturases, the deduced amino acid sequence of pgh9-1 is highly hydrophilic, consistent with the soluble nature of the enzyme. No transmembrane helices were found in the whole polypeptide sequence using an internet program SOSUI (<http://www.tuat.ac.jp/~mitaku/sosui/>).

As indicated in Table 4.1, there is considerable homology between pgh9-1 and $\Delta 9$ stearyl-ACP desaturase cDNAs from other plant species at both nucleotide and deduced amino acid levels. It is a typical observation that the overall amino acid sequence is more conserved than the overall nucleotide sequence in the ORF, which demonstrates the relatively strict conservation at amino acid level. For example, the

sequence similarity between pgh9-1 and *A. thaliana* is 71.4% at nucleotide level, while 84.9% at amino acid level (Table 4.1).

It is also worthy of note that the sequence similarities of the mature peptides between pgh9-1 and other sequences are generally higher than the whole polypeptide. The differences are mainly due to the divergence of the transit peptides. This suggests that the structural requirements for a functional transit peptide is substantially less rigid than those that contribute to catalytic activity.

A stretch of 20 amino acids, located at amino acids 322-342 (underlined in Fig. 4.1), has been found to be a fatty acid desaturase family II signature (<http://www.genome.ad.jp/SIT/MOTIF.html>). The iron-binding motif containing two repeats of EX₂H separated by approximately 100 amino acids found in all Δ^9 stearyl-ACP desaturase and other class II diiron proteins, including the R2 component of ribonucleotide reductase and the soluble bacterial hydrocarbon hydroxylases (Fox et al., 1994) was also present in pgh9-1 (in bold type, Fig. 4.1).

Table 4.1 Sequence Similarities between
pgh9-1 and Other cDNAs for Δ^9 Stearoyl-ACP Desaturase

Plant Species	Accession No.	Similarities (%) to pgh9-1			Reference
		ORF (nt)	Whole Polypeptide	Mature Peptide	
<i>Arabidopsis thaliana</i>	X93461	71.4	84.9	86.3	Piffanelli and Murphy, Unpublished
<i>Brassica napus</i>	X63364	71.8	86.1	88.5	Slocombe et al., 1992
<i>B. rapa</i>	X60978	73.0	86.9	88.8	Knutzon et al., 1992
<i>Elaeis guineensis</i>	U68756	72.4	87.5	89.0	Shah and Rashid, 1996
<i>Sesamum indicum</i>	D42086	74.7	88.9	90.4	Yukawa et al., 1994
<i>Carthamus tinctorius</i>	M61109	72.8	89.2	91.2	Thompson et al., 1991
<i>Coriandrum sativum</i>	M93115	63.3	77.8	78.5	Cahoon et al., 1992
<i>Cucumis sativa</i>	M59858	73.8	88.4	89.9	Shanklin et al., 1991
<i>Glycine max</i>	L34346	74.9	88.8	90.4	Chen and Moon, 1995
<i>Helianthus annuus</i>	U70374	70.8	85.4	88.0	Coughlan et al., unpublished
<i>Linum usitatissimum</i>	X90762	68.4	80.7	82.5	Singh et al., 1994
<i>Asclepias syriaca</i>	U60277	59.2	76.3	76.8	Cahoon et al., 1997
<i>Olea europaea</i>	U58141	69.9	84.7	86.3	Baldoni et al., 1996
<i>Pelargonium xhortorum</i>	U40344	62.7	77.9	79.1	Schultz et al., 1996
<i>Oryza sativa</i>	D38953	67.7	82.1	84.9	Akagi et al., 1995
<i>Ricinus communis</i>	M59857	77.6	89.4	90.7	Shanklin and somerville, 1991
<i>Solanum commersonii</i>	X78935	71.5	87.6	87.6	Tricci et al., unpublished
<i>Simmondsia chinensis</i>	M83199	69.9	86.1	89.0	Sato et al., 1992
<i>Spinacia oleracea</i>	X62898	71.0	86.2	88.2	Nishida et al., 1992
<i>Solanum tuberosum</i>	M91238	70.8	84.8	86.6	Taylor et al., 1992
<i>Thunbergia alata</i> <i>pTAD2</i>	U07597	69.6	86.2	88.5	Cahoon et al., 1994
<i>T. alata pTAD3</i>	U07605	70.5	85.9	87.4	Cahoon et al., 1994
<i>T. alata Δ^6 desaturase</i>	U09269	65.0	75.0	77.8	Cahoon et al., 1994

4.3.2 Gene Expression of pgh9-1 Revealed by Northern Blot Analysis

Total RNAs isolated from developing cottonseed embryos at 25, 30, 36, 45 DAA and leaves, were used for Northern blot analysis. A DNA fragment, based on the 3' UTR of pgh9-1 was used as a gene-specific probe. As demonstrated in Fig. 4.2, high level of expression of pgh9-1 transcript was detected in all four stages of embryo development examined. However, there was no detectable transcript in the leaf tissues. Among the four stages of embryo development, it seems that there was high expression at 30 and 35 DAA, whereas that at 25, 45 DAA were slightly lower. Owing to the difficulty in obtaining RNA samples from very early stage of embryo development, gene expression prior to 25 DAA was not examined. Nevertheless, it can be speculated that the expression of pgh9-1 is induced well before 25 DAA. It is also interesting to observe that at 45 DAA, when the embryos are entering late maturity stage, the expression of pgh9-1 was not drastically reduced.

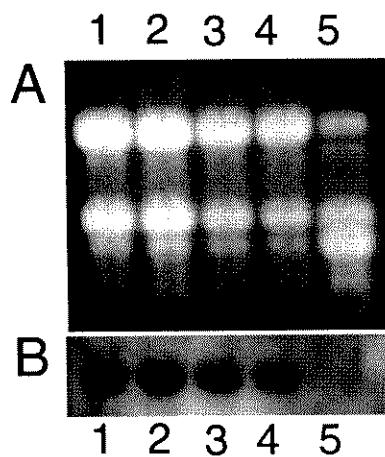


Fig. 4.2 Northern blot analysis of developing embryos and leaf tissues from *G. hirsutum* cv Deltapine-16. (A) EtBr-stained formaldehyde denaturing RNA gel. Lanes 1-4 represent RNA samples isolated from embryos of 25, 30, 36, 45 DAA, respectively. Lane 5 represents RNA isolated from young leaves. (B). RNA gel shown in panel A was transferred to a Hybond N⁺ membrane and hybridised by an [α -³²P]dCTP-labelled pgh9-1-specific probe (3' UTR of pgh9-1).

4.3.3 The Primary Structure of A $\Delta 9$ Stearoyl-ACP Desaturase Gene from *G. hirsutum*

A genomic clone, designated as Gh9-1G, corresponding to the structural gene of pgh9-1 was isolated by screening a *G. hirsutum* genomic DNA library. The DNA sequence determinations indicated that pgh9-1 and the corresponding regions in Gh9-1G are 100% identical. An approximately 1.5 kb section of nucleotide sequence 5' from the translational start site has been determined along with the sequence of three exons and two introns that constitute the entire gene (Fig. 4.3).

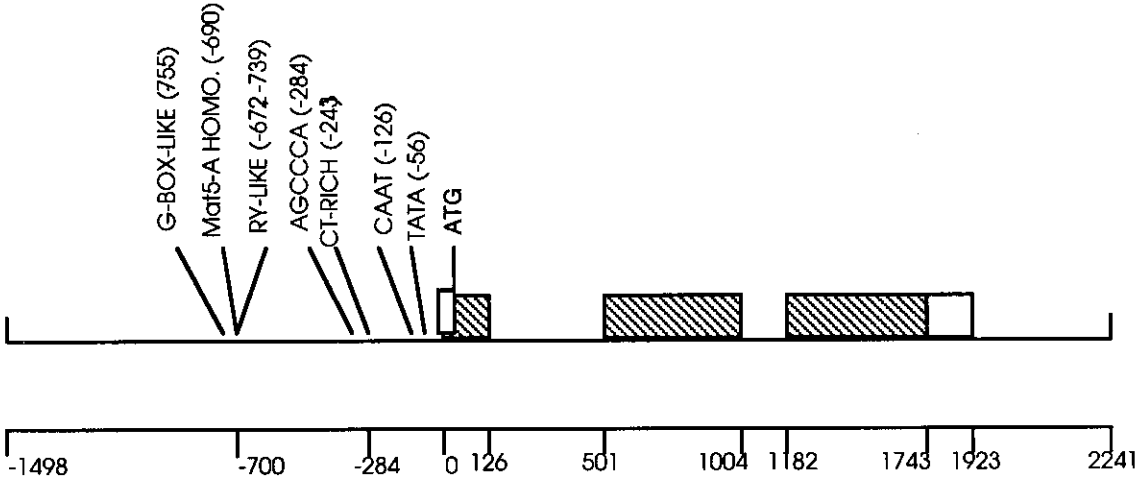
Gh9-1G consists of 3 exons separated by two introns. This is similar to the corresponding gene structures found in *B. napus* (Slocombe et al., 1994) and *S. indicum* (Yukawa et al., unpublished; accession No. D49832). The locations of these two introns within the gene are also conserved among the three plant species, but their lengths are highly variable. As indicated in Table 4.2, the length of intron 1 of Gh9-1G is about 1/3 of *B. napus* and 1/6 of *S. indicum*; intron 2 of Gh9-1G is about twice of *B. napus* but half of *S. indicum*. The high variability of intron size among the three species may indicate that some dramatic internal duplications or deletions of DNA fragments have occurred after their divergence from a common ancestor. Most plant introns are within the size range of 80-139 bp (Simpson and Filipowicz, 1996). In a recent survey of 997 *A. thaliana* introns, only 0.1% that exceeded 1 Kb in length (Brown et al., 1996). Thus it is possible that there is a selection against long introns in plant genes. In this regard, it is notable that the second intron of *B. napus* is only 68 bp. A minimal functional length of 70 bp has been determined for efficient processing of synthetic introns in higher plants (Goodall and Filipowicz, 1990) and consistent with this, naturally occurring introns shorter than 70 bp are rare (Lüehrsen and Walbot, 1994; Simpson et al., 1993). The consensus GT/AG sequence that characterise the 5' and 3' splice sites is present at all the exon/intron boundaries in Gh9-1G.

Fig. 4.3 Gene structure and nucleotide sequence of a $\Delta 9$ stearyl-ACP desaturase gene, Gh9-1G, from *G. hirsutum*.

(A) Organisation of the Gh9-1G. Translated and untranslated exons are indicated by shaded boxes and unshaded boxes, respectively. Introns are indicated by the space between exons. The position of each element is indicated by the distances (bp) relative to the translation site.

(B) Nucleotide and deduced amino acid sequences of Gh9-1G. The translation start site is numbered as 1. Putative regulatory elements, including ABREs-like, RY-like, AGCCCA motifs, CCAAT box, TATA box are indicated in bold type. The putative transcription start site CA, translational start site ATG and the GT/AG nucleotides corresponding to the intron boundaries are also shown in bold. The cotton 2S storage protein Mat5-A homologous region, CT-rich region and the end of transcription are underlined.

A.



B.

TATCTGAAATTTTGTATTATATATTACCTACCTAATAATTATTTTATCTTATTTAAAATA -1438
AATTGATGTATTTATTTCTTTAAATGTGTATGCTAAATTAAAATTAAAGTTTCAAGTATA -1378
TATTTGACCACAATTAGAGTTTCATGTATATAATTACACTAAATTAAAGTTCATATAATT -1318
GCACATTAAATCAAAATTCATATGTAATTTTAAAATTCACCCCTTCTTTAAATATTCAT -1258
TTTATTTTTTATATTCAAACCTGTCTATTATATTCACAACCTGTCACCTGTACATTTAATTT -1198
AAGAGAAAGGATTGTATGAAATAGGGACATTATTACGTCCAATAGTTTGTGTCACATT -1138
AACATTTTCATCCTGAATTTCAAGATTTTCATTTGGATTGTATTTTTCAGGACGAAAT -1078
TGTTGAAATTCAGATAAAAAATAATGATATAATATACAACATTTAAAACAGTGGTGGAT -1018
CTAGGGGACTCGCAGGACCCCGACCCCTAAAATAGAAAATTATTCATTTAAGGCTTTT -958
AAAAATTTTAAAATTTTAAATTAATAAAAAATAAAATTACACTTTAACCTCTTAAAAATAA -898
TAAAAATTTAATTTAATACTTTAAAATTATAAAAAATTAACATTTAAAATTTAAATTT -838
AATTTTAGCTCCCAACATAAATTTTGACTTCACCCCTAATCACACGGTAATTCCTTCTT -778
TTATTCGAAAGAAAGGATGGTCC**CACGTA**ATTAAACTGCC**ATTGCATG**TTTTCTTGGCCACC -718
ABREs-like RY-box like
ACTAGTCTAGTCTACTGC**CACAGTG**TGTGTTTAAAGCTAGAGGTGAT**CATCGG**TTGGGCTAC -658
ABREs-like RY-box like
Mat5-A homologous region
CCGGCCCGACCTGACGGTCCGTCCGAAAAATAAGAGGGTTCGGGTAAAAATATAGGCTCG -598
Mat5-A homologous region
AAATATGGGTTTGAGTAATAAAATGAGGCCCGTTAGAAAACAGGCTGGGCCTCGAGTAAA -538
ACTTTTTTTCTCGGGCCTGGCCCGAATTATGTATTAAATATATATTTTTTTTAAATC -478
AAATATACTTTTTTAATCAAATATATATTTAAATATATACTTTTTTAGTGTTGTAAATTTT -418
AATATGGGTCGGCCCGGGCCGAGCTCGGGCTTAGCAATTTTTTCCGGGTCCGACTTGGAT -358
AAATTTTATAGGCTCATATTTTCGGGCCGGGTGGAATCCGACCTAAAAATAAGCATAAAAT -298
TTTGTCTGGATCC**AGCCCA**AATCTAGCCCGACCCATAATCACCTCTAGTTTAAAGCTTCTT -238
enhancer in soybean β -conglycinin
CTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTTAAAAATTAACATTAAAAATATGTAG -178
CT-rich
AGAAAATCAGCAATTAAAAACAAAAGTTAGGGCTAATGTGTTAAAGTAGCAC**CAATAA**AGT -118
CAAT box
ATCCCTCTCAAGTGAAGTCTTTTCACTTGCAAACAAAAATAATTAAAGACAGAGGAGT -58
CT**TATAA**AGTTAAAGCCGTCCAAAACCCAAACCAGGAAAGGCAAACGAAAGAAAA**ATG** 3
TATA box M
GCTTTGAATTTTAAATGCCATCGCCTCGAAATCTCAGAAGCTCCCTTGCTTTGCTCTTCCA 63
A L N F N A I A S K S Q K L P C F A L P
CCAAAGGCCACCCTTAGATCTCCCAAGTTTTCATGATCTCCACCATTCTTCTGGCTCC 123
P K A T L R S P K F S M I S T I P S G S

Fig. 4.3B Con't

[illegible]

The other distinguishable feature of these two introns of Gh9-1G is their AT-rich nature. As demonstrated in Table 4.2, while the ORF contains 54% AT, the AT content of the two introns is 69 and 70%, respectively. Similar observations can also be made for the corresponding gene in *B. napus* and *S. indicum*, as summarised in Table 4.2. In fact, AT-richness is a common feature of plant pre-mRNA introns and it is also suggested that AT-richness of introns is essential for their recognition by the splicing machinery and efficient splicing (Simpson and Filipowicz, 1996).

Table 4.2 The AT Contents of Intron vs. ORF in $\Delta 9$ Stearoyl-ACP Desaturase Gene in Three Plant Species (*B. napus*, *S. indicum* and *G. hirsutum*)

		<i>B. napus</i>	<i>S. indicum</i>	<i>G. hirsutum</i>
1st intron	length	1295	2417	374
	AT content (%)	67%	64%	69%
2nd intron	length	68	321	178
	AT content (%)	65%	63%	70%
ORF	length	1200	1191	1194
	AT content (%)	50%	53%	54%

A putative TATA box (TATAAA) is present 56 bp upstream from the translational start codon, resembling the plant TATA box consensus sequence of (C/G)TATA(T/A)(A₁₋₃)(C/T)A (Messing et al., 1983). Although the exact transcription initiation site has not been mapped, near the required 30 bp downstream from the beginning of the TATA

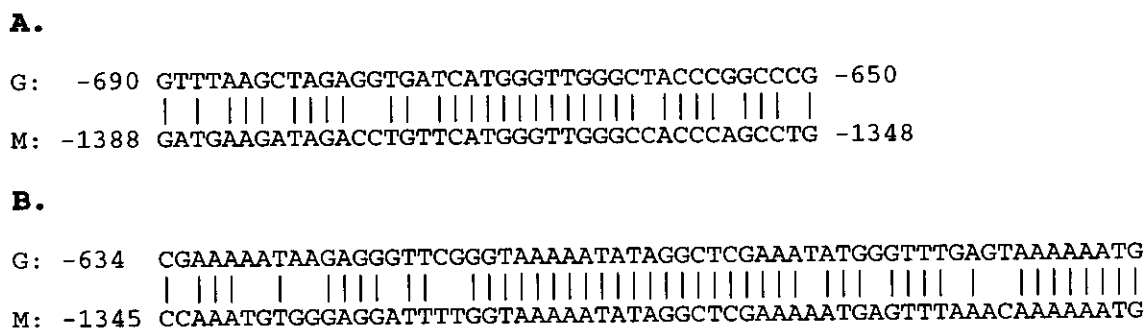


Fig. 4.4 Gap analyses of two homologous regions (A and B) in the promoter of Gh9-1G and those of cotton 2S albumin storage protein (Mat5-A) gene. Their position in respective gene are numbered. A region: Identities = 32/41 (78%)
 B region: Identity = 48/61 (78%). G: G9-1G; M: Mat5-A.

4.3.4 The $\Delta 9$ Stearoyl-ACP Desaturase Gene Family

Southern blot analyses were carried out at high stringency with DNA probes consisting of either a pgh9-1-specific DNA fragment corresponding to the 3' UTR of the gene or a fragment covering the entire coding region of pgh9-1. Three different enzymes, EcoRI, HindIII and XbaI which do not cut in the region of probe DNA sequences, were used to digest genomic DNAs from *G. barbadense*, *G. hirsutum* (cv Deltapine-16), *G. hirsutum* (cv Siokra), *G. herbaceum*, *G. raimondii*. As shown in Fig. 4.5-A, when the genomic DNA Southern blot was probed by a pgh9-1-specific probe (3' UTR), among all three enzyme restriction patterns, two strong bands were detected in each of the allotetraploids whereas a single strong band was found in each of the two diploids. Under the circumstances that each band contains only one DNA fragment, the restriction fragment pattern shown in Fig. 4.5-A probably indicates that there are two copies of pgh9-1 in each of the tetraploids and only one copy in each of the diploids examined, consistent with their ploidy levels. However, when the coding region probe was used, it revealed a complex pattern consisting of multiple fragments in each of the three enzyme digests (Fig. 4.5B). This clearly indicates that the $\Delta 9$ stearoyl-ACP

desaturase is encoded by a multigene family which is composed of 6-8 genes (or copies) in diploid cotton genomes, and probably twice as many in the allotetraploid cottons (*G. barbadense* and *G. hirsutum*). Some bands appear to contain individual fragments, while others may contain multiple fragments as indicated by their obviously stronger signals. Of course, the uneven band intensity on the same lane might also indicate the variable homologies between the gene represented by the band and the probe DNA.

The allotetraploid cottons, *G. barbadense* and two *G. hirsutum* cultivars share highly conserved restriction fragment length patterns, with only minor differences. However, the two diploid species are distinctively different. It was estimated that the A- and D-genomes diverged from a common ancestor 6-11 million years ago, whereas the divergence of the two tetraploids occurred much more recently, about 1-2 million years ago (Endrizzi et al., 1985; Wendel and Albert, 1992). Among the currently existing diploid *Gossypium* species, *G. herbaceum* and *G. raimondii* are regarded as most close to the progenitor species of the modern tetraploid cotton (Endrizzi et al., 1985). Indeed, in Fig. 4.5A, there are the two EcoRI bands in each of the three tetraploids (lanes 1-3), and the lower band corresponds to the A-genome diploid *G. herbaceum* (lane 4), while the upper band corresponds to the D-genome diploid *G. raimondii* (lane 5). Even among the complex patterns in Fig. 4.5B, some bands in tetraploids are identical to those of diploids. It should also be possible to assign the putative subgenomic origins of some tetraploid bands by comparing with the A- and D-genome diploids in Fig. 4.5B.

It is noteworthy that the D-genome lane always has the strongest bands relative to other genomes. This is due to the fact that the size of D-genome is the smallest, being 65% of the A-genome and only 39% of the tetraploids (Endrizzi et al., 1985). Therefore, when the same amount of DNA is applied on each lane, the D-genome lanes always have the strongest bands.

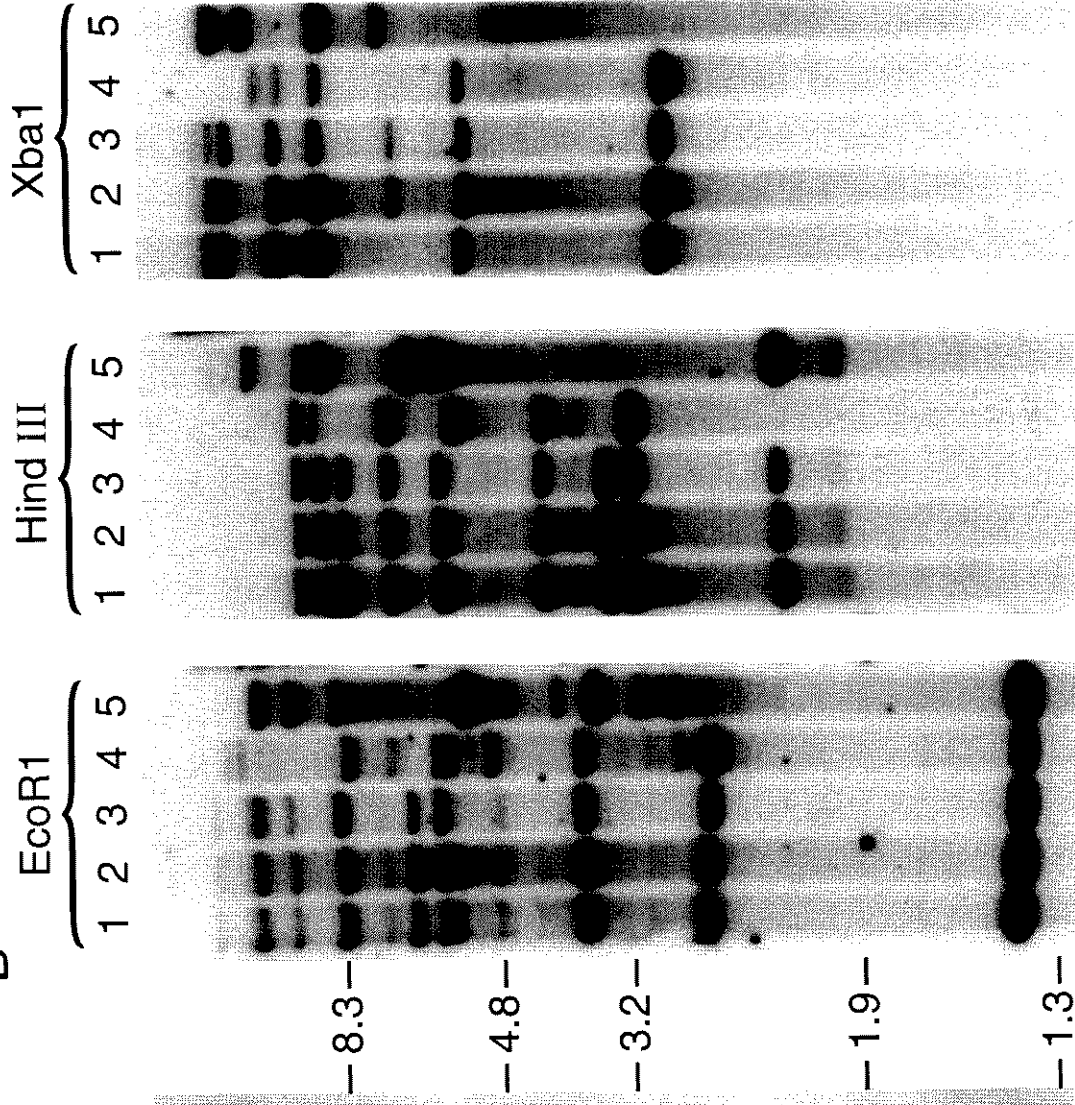
Fig. 4.5 Southern blot analysis of $\Delta 9$ stearyl-ACP desaturase genes in *Gossypium* species.

(A) probed by pgh9-1-specific probe (3' UTR of pgh9-1)

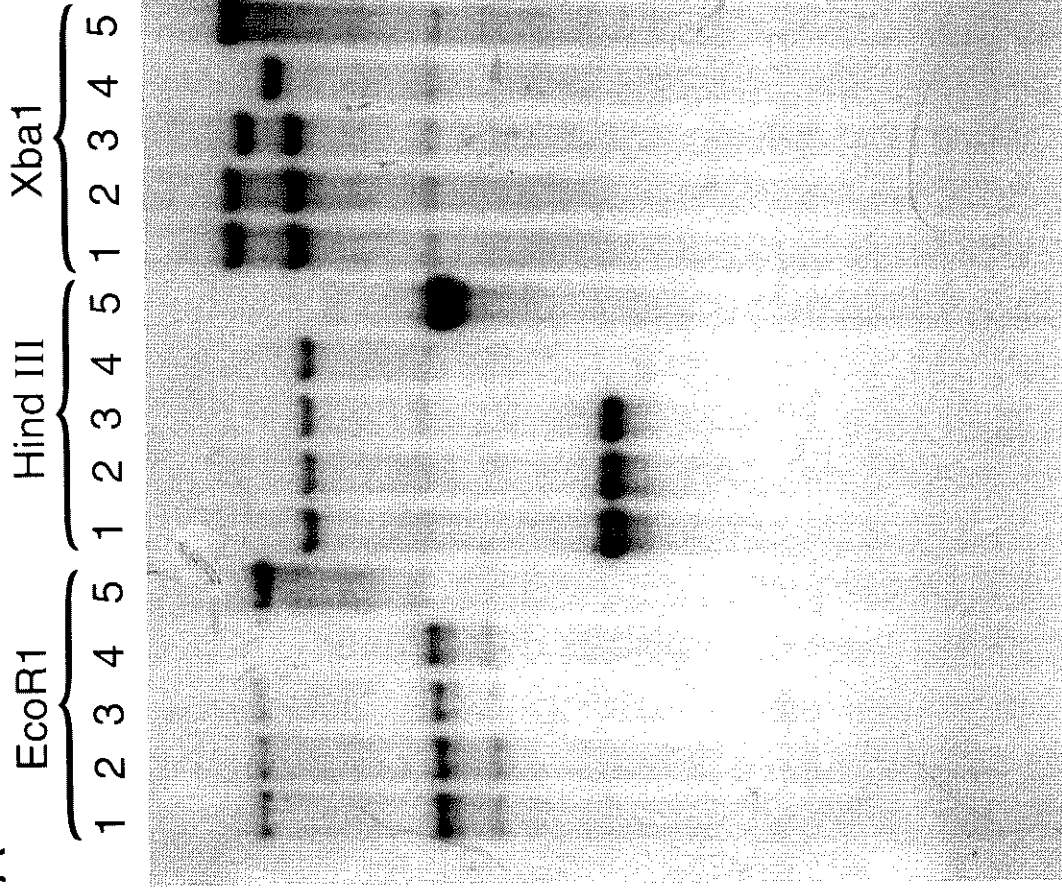
(B) probed by the coding region of pgh9-1.

- lane 1: *G. barbadense*
- lane 2: *G. hirsutum* cv Deltapine-16
- lane 3: *G. hirsutum* cv Siokra
- lane 4: *G. herbaceum*
- lane 5: *G. raimondii*

B



A



4.4 DISCUSSION

In this chapter the isolation and characterisation of a cDNA encoding the $\Delta 9$ stearoyl-ACP desaturase (pgh9-1) and its corresponding structural gene (Gh9-1G) in *G. hirsutum* is described. Similar to other plant $\Delta 9$ stearoyl-ACP desaturase cDNAs, pgh9-1 encodes a polypeptide of 397 amino acids with the first 32 amino acid residues being a transit peptide. A pair of short sequences, EX₂H (ENRH and EKRH), separated by about 100 amino acid residues in the deduced amino acid sequence of pgh9-1 was present. Although lacking sequence homology with $\Delta 9$ desaturase genes from mammals, fungi and bacteria, pgh9-1 shares reasonably high homology with other plant acyl-ACP desaturases. To summarise the nucleotide similarities among these genes, the nucleotide sequences in the entire ORFs of 25 different acyl-ACP desaturase cDNAs were used for a sequence comparison by a PILEUP analysis (GCG package) and the results are shown as a dendrogram in Fig. 4.6.

As shown in Fig. 4.6, the cDNAs encoding various acyl-ACP desaturases other than a $\Delta 9$ stearoyl-ACP desaturase have been classified into a group, with the exception of acyl-ACP desaturase from *B. napus* which is highly homologous to the $\Delta 9$ stearoyl-ACP desaturase from *B. rapa*. Among the cDNAs for $\Delta 9$ stearoyl-ACP desaturases, the two monocots, rice and oil palm form a distinct clade separated from the rest which all represent dicot genes. The flax sequence is the only one not included in the dicot group. Among the dicots, some gene sequences are clustering according to their respective phylogenies. For example, the two *Solanum* (*S. commersonii* and *S. tuberosum*) sequences are in one clade and the four *Brassicaceae* genes (*A. thaliana*, *B. napus* and *B. rapa*) in one group. The two *Asteraceae* sequences, *H. annuus* and *C. tinctorius* are in one group and two *T. alata* sequences are another group. This is not a phylogenetic tree, therefore it is not surprising that some sequences are not grouped reflecting the phylogenetic position of the species they represent. For example, both *O. europaea* (*Oleaceae*) and *S. indicum* (*Pedaliaceae*) belong to *Scrophulariales*, but they are not grouped together. Rather, *O. europaea* is more similar to *T. alata*, whereas *S. indicum* is

close to *G. hirsutum* and *R. communis*, both are *Malvaceae*. *S. chinensis*, together with *R. communis*, are classified to *Euphorbiales* in *Malvaceae*, but are located far apart in the present dendrogram.

The relationships among these acyl-ACP desaturase genes was also studied by both distance-based and parsimony analysis to produce phylogenetic trees of the plant $\Delta 9$ stearoyl-ACP desaturases based on their DNA and polypeptide sequences. However, none of these analysis turned out to be closely related to their true phylogenetic identities in the higher plant kingdom (data not presented). Therefore, the $\Delta 9$ stearoyl-ACP desaturase gene sequences may not be suitable for phylogenetic studies among higher plants.

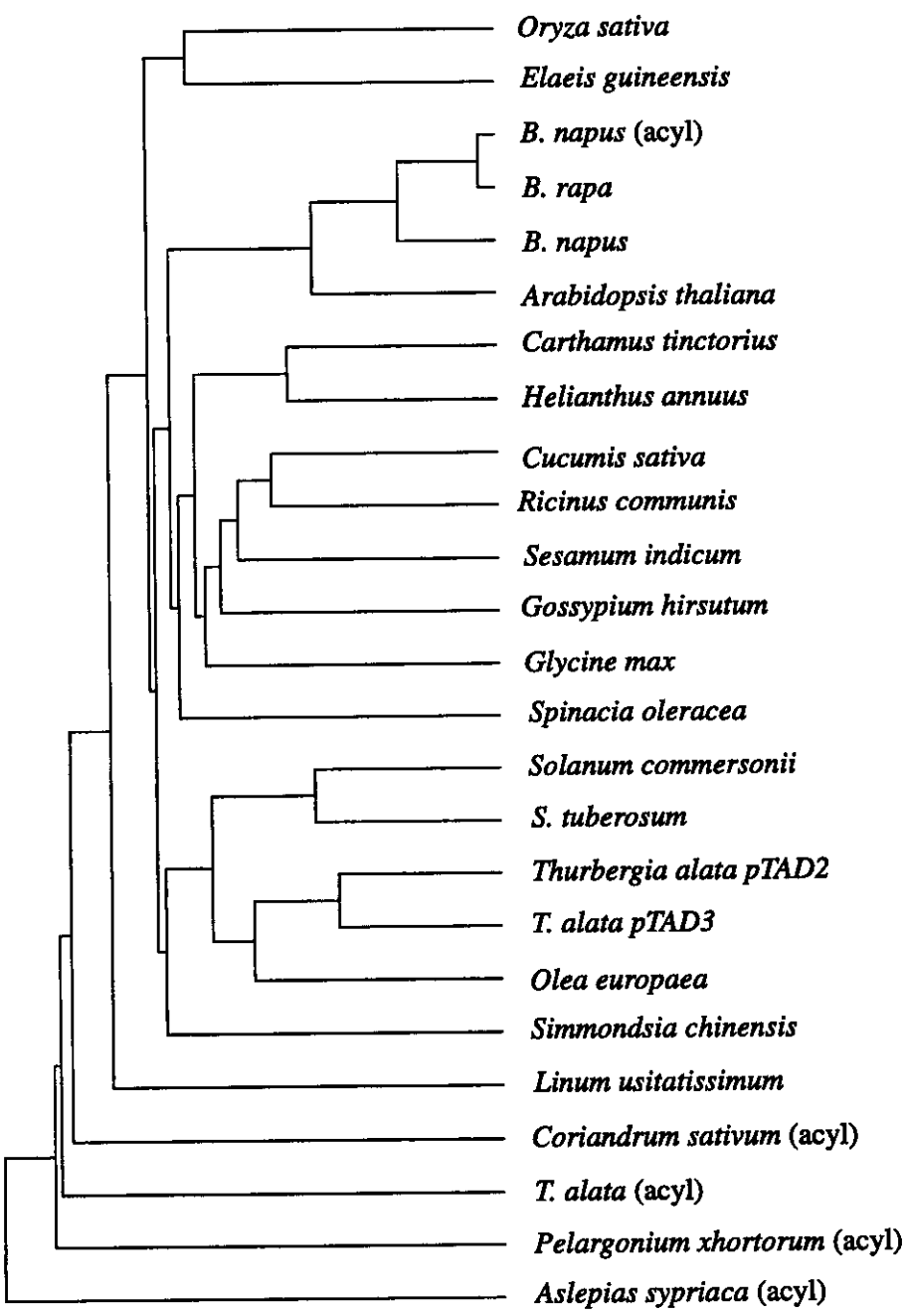


Fig. 4.6 A dendrogram showing the relationship between various acyl-ACP desaturases according to their nucleotide sequence similarities in the ORF. It was drawn by a GCG program PILEUP.

Northern blot analysis using the pgh9-1-specific probe indicated that the expression of this gene was highly induced during cottonseed embryo development, but no expression was observed in leaf tissues. Therefore, it is presumed to be responsible for the introduction of first double bond into stearyl-ACP during cottonseed fatty acid synthesis and oil accumulation in cottonseeds.

There are three major seed storage products, i.e. oil, storage protein and oleosin in most plant seeds and they accumulate at different points during embryo development. It is generally believed that the onset of oil synthesis occurs during the first half of embryo development and precedes that of storage proteins (Goldberg et al., 1994). In the current experiment, it was found mRNA transcripts of pgh9-1 were relatively abundant during all periods of seed development examined, ranging from 25 to 45 DAA. Because of the difficulty in obtaining sufficient material for Northern blot analysis in the very early stage of cotton embryo development, the expression of pgh9-1 earlier than 25 DAA was not examined. Therefore, alternative techniques, such as *in situ* hybridisation, or the use of materials from *in vitro* somatic embryogenesis are required in the future in order to obtain a more detailed information on the temporal and spatial gene expression patterns of $\Delta 9$ stearyl-ACP desaturase in cottonseed.

Comparative studies on the temporal and spatial expression of oleosin and $\Delta 9$ stearyl-ACP desaturase genes has been reported from *B. napus* by *in situ* hybridisation at five different stages of development ranging from the torpedo stage to a mature-desiccating embryo (Slocombe et al., 1994). The temporal pattern of gene expression varied dramatically between the two classes of genes. The expression of $\Delta 9$ stearyl-ACP desaturase gene was relatively high, even at the torpedo stage, whereas oleosin gene expression was barely detectable at this stage. By the stage of maximum embryo fresh weight, the expression of $\Delta 9$ stearyl-ACP desaturase gene has declined considerably while oleosin gene expression was at its height.

The cotton $\Delta 9$ stearyl-ACP desaturase gene, pgh9-1, is a member of a multi-gene family comprising of at least 6-8 genes (or copies) per diploid genome in *Gossypium*. The various members of the $\Delta 9$ stearyl-ACP desaturase (or closely related genes) gene family may share high homologies, as indicated by the strong signals obtained in the Southern blot analysis with the coding region of pgh9-1 under high stringency conditions. However, the Southern blot analysis with the gene specific probe indicated that each of the diploids (*G. herbaceum* and *G. raimondii*) contains one DNA restriction band harbouring at least one gene copy, and each of the allotetraploids (*G. hirsutum* and *G. barbadense*) contains two bands, each of which represents its subgenomic origin.

The fact that $\Delta 9$ stearyl-ACP desaturase is encoded by a multi-gene family is one of the most puzzling aspects in the studies of plant fatty acid desaturase genes. Even within *A. thaliana* with its small genome, the $\Delta 9$ stearyl-ACP desaturase is encoded by at least five genes (Baerson and Lamppa, 1993; Hlousek-Radojcic et al., 1992). In contrast, *B. napus*, an allotetraploid species, contains only four closely related $\Delta 9$ stearyl-ACP desaturase genes with two originating from each of the parent species *B. rapa* and *B. oleracea* (Slocombe et al., 1994). It has been considered to be advantageous to have multiple genes to allow fine tuning of expression in different tissues or to function at extreme environmental conditions. It will be of great interest in future studies to explore the functional aspects of other $\Delta 9$ stearyl-ACP desaturase genes in cotton.

As is the case in *B. napus* and *S. indicum*, the structural gene of *G. hirsutum* $\Delta 9$ stearyl-ACP desaturase, Gh9-1G, consists of three exons and two introns. Although the intron sequences and lengths are highly variable among these three species, their positions within the gene are relatively conserved. This supports the view that the $\Delta 9$ stearyl-ACP desaturase genes in higher plants share a common origin. Several previously identified seed-specific promoter motifs, including ABREs-like, RY-like and a AGCCCG motif, were also found in the 5' flanking region of the Gh9-1G. The

ABREs elements are known to be involved in the ABA-dependent induction of many genes and have been shown to bind proteins belonging to the bZIP family of transcription factors (Giraudat et al., 1994).

Although the ACGT core sequence has been suggested to be of importance for high affinity binding of ABREs-binding proteins (Izawa et al., 1993), it has been shown that *cis*-elements with similar but not identical sequences can also be the target for the ABREs-binding transcriptional complexes (Schindler et al., 1992; Pysh et al., 1993). Furthermore, the transcription factor *Opaque2*, which is a bZIP protein, binds with high affinity to a sequence lacking an ACGT core (de Pater et al., 1994). The core motif has been found in most cotton *Lea* and *LeaA* genes (Galau et al., 1992a), the cotton oleosin gene (Hughes et al., 1993) and cotton storage protein genes (Galau et al., 1991; 1992b). ABA is important in maintaining embryogenic development during the storage reserve accumulation phase of embryogeny, and in mediating suppression of precocious germination (Goldberg et al., 1989; Thomas et al., 1991; Thomas, 1993). Microspore-derived haploid embryoids from *B. napus* increase fatty acid production, especially of long-chain forms, when exposed to ABA (Holbrook et al., 1992). Production of linoleic acid metabolites in very young maize embryos is inducible by ABA (Abián et al., 1991). Among the genes involved in the fatty acid synthesis in higher plants, the ACP promoter from *B. napus* contains the core of an ABREs motif. A single ABREs motif is also present in the promoter of $\Delta 9$ stearoyl-ACP desaturase gene of *B. napus*, and its position in the gene (at the -774 relative to transcription start) is similar to the two ABREs-like motifs found in Gh9-1G. In *B. napus*, ABA was shown to increase the levels of transcripts for $\Delta 9$ stearoyl-ACP desaturase in zygotic embryos *in vitro* (Slocombe et al., 1994).

The CATGCATG or RY motif is widely distributed in dicot and monocot genes (Bäumlein et al., 1986; Dickinson et al., 1988). Deletion of the RY repeat in the *Vicia faba* storage protein gene LeB4 and the soybean glycinin gene resulted in a decrease in

reporter gene expression in transgenic tobacco seeds (Bäumlein et al., 1994; Lelievre et al., 1992).

The AGCCCA motif at position -277 (in Gh9-1G) was shown to act as an enhancer of seed-specific expression in soybean b-conglycinin (Chen et al., 1988). The position of this element is also relatively well conserved between the two genes (-280 in soybean b-conglycinin).

The CT-rich region found in the Gh9-1G promoter also occurs in the immediate upstream of TATA-like box in the promoter of $\Delta 9$ stearyl-ACP desaturase gene in *B. napus* (Slocombe et al., 1994). In fact, almost all the characterised promoters or 5' UTR of genes involved in the pathway of fatty acid synthesis possess a similar CT rich region or (CT)_n tract. These include *A. thaliana* biotin carboxylase (Bao et al., 1997), in the 5' UTR sequences of biotin carboxylase from tobacco (Shorrosh et al., 1995), three ACP genes from *A. thaliana* (Lamppa and Jacks, 1991; Post-Beittenmiller et al., 1992), two ACP genes from *B. napus* (de Silva et al., 1990), $\Delta 9$ stearyl-ACP desaturase gene from *B. napus* (Slocombe et al., 1994), 1-acyl-sn-glycerol-3-phosphate acyltransferase gene (Knutzon et al., 1995). Previous studies have suggested that the d(CT)_n tracts in eukaryotic genomes may be involved in recombination (Weinreb et al., 1990), replication (Caddle et al., 1990), transcription (Wells et al., 1988), and chromatin structure (Lu et al., 1993). Its conservation in the genes involved in fatty acid synthesis suggests that it may have a universal regulatory function.

AT-richness in the distal region of the Gh9-1G promoter is a general characteristic of most promoter sequences. However, it has also been suggested that the AT-rich motifs may bind to ubiquitous, prevalent nuclear proteins and are found in the promoters of most, if not all, seed protein genes (Thomas, 1993). Hence it is possible that they could be acting as nonspecific enhancers by functioning as scaffold attachment regions creating transcriptionally active domains. For example, phaseolin UAS3 contains an A/T-rich domain that acts as an enhancer in transgenic tobacco (Bustos et al., 1989; 1991).

Perhaps the most interesting finding about the Gh9-1G promoter sequence is a region homologous to the cotton 2S storage protein Mat5-A promoter (Galau et al., 1992b). It is located near the ABREs-like and RY-like motifs in Gh9-1G. This region may contain regulatory elements which can enhance or refine the seed specificity. This aspect of this region needs to be studied in future deletion analysis and site directed mutagenesis.

As shown in Fig. 4.4A, the putative regulatory elements are located at two regions separated by more than 400 bp. Therefore, the promoter structure of Gh9-1G demonstrates a bipartite organisation similar to the model established in seed protein and *Lea* class gene promoters (Thomas et al., 1991). In this model, proximal promoter regions specify seed expression and more distal regions function to refine and enhance the basic expression patterns conferred by proximal regions. This bipartite model was developed for seed protein and *Lea* class genes. Whether it also applies to the $\Delta 9$ stearoyl-ACP desaturase in cotton or other highly expressed, tissue-specific genes in higher plants remains to be examined by further studies.

In conclusion, a $\Delta 9$ stearoyl-ACP desaturase gene with high expression in developing cotton embryos has been cloned. The availability of such a gene should permit the genetic manipulation of fatty acid desaturation in cottonseed oil through antisense or co-suppression techniques. Constructs with both sense and antisense genes driven by seed-specific promoter have been made (section 2.10 in Chapter Two) and future experiments with cotton gene transformation will undoubtedly contribute to the understanding of how much this gene contributes to the fatty acid desaturation in cottonseed oil. The unveiling of the genomic structure and some potentially useful putative regulatory elements of the $\Delta 9$ stearoyl-ACP desaturase gene in cotton also provides some vital information which will be the basis for further studies on the regulation of gene expression. This is particularly important due to the observed complexity of the cotton $\Delta 9$ stearoyl-ACP desaturase multigene family as revealed in this current study.

Chapter Five

Isolation and Characterisation of cDNA Clones Encoding Microsomal ω -6 Fatty Acid Desaturase in *G. hirsutum*

5.1 INTRODUCTION

In plants, the initial desaturation reaction is carried out in plastids by the soluble plastidial Δ^9 stearoyl-ACP desaturase (McKeon and Stumpf, 1982). Further desaturation can then occur either in the plastids or on the endoplasmic reticulum (microsomes) by two parallel fatty acid desaturation pathways (Somerville and Browse, 1991). The first extra-plastidial desaturation in plants is catalysed by a microsomal ω -6 desaturase enzyme. This enzyme converts oleic acid esterified on the sn-2 position of phosphatidylcholine to linoleic acid with the reduction of molecular oxygen to water, and requires the presence of NADH and an electron transport chain that consists of cytochrome b_5 and NADH:cytochrome b_5 reductase (Smith et al., 1990; Kearns et al., 1991). It has also been established that the microsomal ω -6 desaturase catalysed pathway is the primary route of production of polyunsaturated seed storage lipids (Heinz, 1993).

Cottonseed oil contains more than 50% polyunsaturated fatty acid, mainly linoleic acid, and is usually chemically hydrogenated to increase its shelf-life and stability during

frying. Production of margarines and shortenings from cottonseed oil also involves extensive chemical hydrogenation. The extra industrial process not only increases the oil cost, but also results in the formation of *trans* isomers of monounsaturated fatty acids. These isomers have long been under suspicion for their possible association with coronary heart disease in human (Khosla and Hayes, 1996). Hence it is desirable to breed a novel cottonseed genotype with its oil high in monounsaturates at the expense of polyunsaturates. The recent progress in the gene cloning and the development of antisense or co-suppression technology provides an attractive strategy for modifying the degree of desaturation of fatty acids in seed oil of many oilseed crops including cottonseed.

Although the microsomal ω -6 desaturase has not been purified, a locus corresponding to its structural gene was first isolated from *A. thaliana* by T-DNA tagging (Okuley et al., 1994). There has been great interest in studying the structure and function relationships of this desaturase from a number of sources, attempting to understand the desaturation mechanism and basis for their substrate and positional specificities. Furthermore, the availability of cloned genes native to the target plant species is a prerequisite for an efficient genetic engineering by means of the specific expression of sense or antisense mRNA which inhibits the production of selected desaturases. The efficiency of such a genetic engineering approach depends also on several other factors, one of which is the number of copies of the endogenous target gene. In this chapter, I describe the isolation of two different ω -6 microsomal genes in *G. hirsutum*, and present evidence that microsomal ω -6 desaturase gene in cotton is encoded by a small gene family, while pghD12-1 may play a major role in converting oleic acid into linoleic acid during the biosynthesis of cottonseed oil.

5.2 MATERIALS AND METHODS

5.2.1 Isolation of cDNA Clones from a Cottonseed λ cDNA Library

To isolate the cDNAs encoding microsomal ω -6 desaturase from cotton, a heterologous probe consisting of the entire coding region of microsomal ω -6 desaturase from *B. juncea* was used to screen a cottonseed cDNA library which was constructed as described in Chapter Two. Approx. 1×10^6 Pfu were screened and subsequently, twelve positive plaques identified in the primary screening were purified by two further consecutive screenings and their cDNA inserts in pBluescript vector were characterised by nucleotide sequence determination. DNA sequencing was performed with an ABI 373 system (Applied Biosystems) using dye terminator and dye primer sequencing reactions. Sequence comparisons and alignments were performed with programs in the GCG package (Devereux et al., 1984).

5.2.2 Amplification of the cDNA Ends from λ cDNA Library DNA

To obtain the missing 5' end of cDNA of pghD12-1, a PCR approach using crude DNA extracts from λ cDNA library as DNA template was employed. A sample of λ library phage containing approx. 1×10^9 Pfu was extracted by phenol/chloroform and chloroform followed by ethanol precipitation. The pellet was air-dried and dissolved in 20 μ l H₂O. A cDNA-specific primer, *d12A4* (5'- GCATAGGTCATGGACCACGT -3') was used in combinations with either the forward (5'- GTAAAACGACGGCCAGT -3') or reverse (5'- GGAAACAGCTATGACCATG -3') pBluescript vector-specific primers that bind close to the 5' and 3' EcoRI cloning site. A 50 μ l PCR reaction mixture contained 200 μ M dNTP, 1x PCR buffer, 20 pmol of each of the primers, 5 μ l of cDNA library extracts and 1 unit of Taq polymerase. PCR was conducted with an initial denaturation step of 94°C for 2 min followed by 30 cycles at 94°C for 45 seconds, 57°C for 1 min, and 72°C for 1 min. After the last cycle, PCRs were extended for 10 min at 72°C. The PCR product was purified by Wizard PCR Preps DNA Purification System

(Promega) and cloned into T-vector (Promega) according to manufacturer's instructions.

5.2.3 Primer-adapter Mediated PCR

The protocol of primer-adapter mediated PCR described by Espelund and Jakobsen (1992) was simplified and used to amplify the 5' upstream DNA fragment using genomic DNA as templates. Ten microgram genomic DNA isolated from freeze-dried leaves of cotton (*G. hirsutum* L. cv Deltapine-16) was digested by BglII overnight. Adapter with the BglII overhangs was ligated to the restricted cotton genomic DNA in a total 20 µl volume including 1x ligation buffer (Gibco BRL/Life Technologies), 200 ng adaptors, 5 µg genomic DNA and 1 unit T4 DNA ligase. The sequences of the BglII adapter are as follows: top strand: 5'- GCACTCTCCAGCCTCTCACCGCAG -3'; bottom strand: 5'- AGATCTGCGGTGAGAGGCTGGAGAGTGC -3'. The ligation reaction was incubated at 16°C overnight, and then ethanol precipitated and washed in 70% ethanol twice prior to being air-dried and dissolved in 50 µl H₂O. A primer based on the known cDNA sequence of pghD12-2, *3111a1* (5'- CTGAAAGCATGGTGGCC-ACATTC -3'), in combination with BglII adapter-primer (the top strand of the adaptors) was used in the following PCR. A hot start PCR was performed with an initial cycle at 95°C for 4 min before the Taq polymerase was added. This was followed by 30 cycles at 94°C for 1 min, 60°C for 1 min and 72°C for 1 min. The PCR reaction was terminated with a 10 min extension at 72°C. The procedures for the PCR product purification, TA cloning and DNA sequencing were the same as described in above section.

5.2.4 RNA Northern Blot Analysis

Total RNA was extracted from young leaves and immature embryos (25, 30, 36, 45 DAA) of *G. hirsutum* cv Deltapine-16 using a protocol described in section 2.4.5 in

Chapter Two. The RNA gel electrophoresis and the following Northern blot hybridisation was carried out at standard conditions as described in section 2.7.

5.2.5 Genomic DNA Southern Blot Analysis

Genomic DNAs were prepared from young leaves of *G. barbadense*, *G. hirsutum* (cvs Deltapine-16 and Siokra), *G. arboreum*, *G. herbaceum*, *G. raimondii*, *G. trilobum* and *G. robinsonii* and further purification by CsCl gradient was carried out as described in Chapter Two. Ten microgram DNA were digested with EcoRI or HindIII or XbaI and electrophoresed through a 0.7% agarose gel. The gel was blotted onto a Hybond-N⁺ nylon membrane (Amersham). The filters were probed with [α -³²P]dCTP-labelled DNA fragments corresponding to either the coding region of pghD12-1 or the gene-specific fragments based on the 3' UTR sequences of cDNAs, pghD12-1 or pghD12-2, respectively. Hybridisations were carried out under the standard conditions as described in section 2.7 in Chapter Two. Following the hybridisation, the filters were at moderate stringency conditions with an initial wash in 2x SSC, 0.1% SDS at 55°C, followed by two washes in 0.5x SSC, 0.1% SDS at 55°C prior to autoradiography.

5.3 RESULTS

5.3.1 Isolation and Sequence Analysis of cDNAs Encoding Microsomal ω -6 Desaturase from *G. hirsutum*

5.3.1.1 Isolation of two 5' truncated cDNAs from cDNA library

A cottonseed cDNA library was screened using a DNA fragment consisting of the coding region of the microsomal ω -6 desaturase from *B. juncea* as a probe. Twelve strongly hybridising plaques obtained in the primary screening were purified by two further consecutive screenings. Restriction enzyme mapping and partial DNA sequence analysis indicated that the excised plasmids from eleven of them harboured the same insert which was named as pghD12-1; while the twelfth had an insert which showed

considerable homology to pghD12-1 in the coding region, but an unique 3' UTR. This clone was named as pghD12-2. Analyses of the nucleotide sequences of pghD12-1 and pghD12-2 revealed that both of them were truncated at the 5' ends. Clone pghD12-1 was 1248 bp long, lacking an estimated 84 bp; while pghD12-2 was 1131 bp long, lacking an estimated 195 bp from the 5' end of the ORF. These estimations were made by comparison with the microsomal ω -6 desaturase in *A. thaliana* and *B. juncea*.

5.3.1.2 PCR amplification of 5' end of pghD12-1 from cDNA library

A PCR approach was successfully adopted to obtain the 5' ends of pghD12-1 from the crude DNA extracts of a cottonseed λ cDNA library. PCRs with *d12A4* primer and either forward or reverse primer on the pBluescript vector yielded products of expected size. One cDNA end was found to match perfectly with the overlapping region of the original cDNA clone and therefore it was determined to be the missing 5' end of the pghD12-1. The complete pghD12-1 is 1411 bp in length. It contains an ORF encoding a peptide of 386 amino acids, 78 bp of 5'UTR and 175 bp 3' UTR (Fig. 5.1).

```

CTCGCCCAAAACCAACACGCCTTCTTTGCCTCGTGTTCATCACCTGGCGTTAAACTGCT 60
TTCTTTTAAAGCCAGCAAAATGGGTGCCGGTGGTAGGATGCCAATTGACGGTATAAAGGAG 120
      M G A G G R M P I D G I K E
GAAAATCGAGGCTCGGTCAATCGAGTTCCGATCGAGAAGCCTCCGTTTACGCTCGGTCAG 180
E N R G S V N R V P I E K P P F T L G Q
ATCAAGCAAGCCATTCCGCCCCACTGTTTTCGCCGCTCCCTCCTTCGATCCTTCTCCTAC 240
I K Q A I P P H C F R R S L L R S F S Y
GTGGTCCATGACCTATGCTTAGCCTCTTTCTTTTACTACATTGCAACATCATATTTTCAC 300
    d12A4 primer
V V H D L C L A S F F Y Y I A T S Y F H
TTTCTCCCACAACCCCTTTTCCTACATTGCTTGGCCTGTCTATTGGGTTCCTCAAGGTTC 360
F L P Q P F S Y I A W P V Y W V L Q G C
ATCCTCACCGGTGTTTGGGTTCATCGCACACGAGTGGGGTCACCACGCTTTCAGAGACTAC 420
I L T G V W V I A H E W G H H A F R D Y
CAATGGGTGACGACACCGTCCGGTTGATCCTTCATTCCGCCCTTTTAGTCCCGTACTTC 480
Q W V D D T V G L I L H S A L L V P Y F
TCGTGGAAAATCAGTCACCGCCGTCACCACTCGAACACCGGTTCCATGGAGCGTGACGAA 540
S W K I S H R R H H S N T G S M E R D E
GTATTCGTGCCCAAAACCAAGTCTAAATTATCATGCTTTGCGAAATACTTAAACAATCCA 600
V F V P K P K S K L S C F A K Y L N N P
CCCGGTGCGAGTTCATCTCTTGTAGTCACATTGACTCTTGGTTGGCCTATGTACTTAGCC 660
P G R V L S L V V T L T L G W P M Y L A
TTCAACGTTTCGGGTGCGATACTATGATCGATTAGCTTCCCACTATAACCCCTTATGGCCCC 720
F N V S G R Y Y D R L A S H Y N P Y G P
ATTTACTCCGATCGCGAGAGGCTACAAGTTTACATCTCCGATACTGGTATATTTGCGGTA 780
I Y S D R E R L Q V Y I S D T G I F A V
ATTTATGTACTTTTATAAGATTGCTGCAACAAAAGGGCTGGCTTGGCTTTTATGCACCTAT 840
I Y V L Y K I A A T K G L A W L L C T Y
GGGGTGCCCTCTACTTATTGTGAATGCCTTCCTTGTGTTGATCACCTACTTGCAACATACT 900
G V P L L I V N A F L V L I T Y L Q H T
CACTCGGCATTGCCGCATTATGACTCGTCCGAATGGGATTGGTTGCGAGGAGCATTTGTCG 960
H S A L P H Y D S S E W D W L R G A L S
ACGATGGATCGAGATTTCCGGGTGTTGAACAAAGTGTTCCATAACATCACCGATACGCAT 1020
T M D R D F G V L N K V F H N I T D T H
GTTGCTCATCACCTCTTCTCAACGATGCCACATTATCATGCAATGGAGGCCACTAAAGCA 1080
V A H H L F S T M P H Y H A M E A T K A
ATCAAACCAATACTCGGCAAGTATTATCCTTTTCGACGGGACACCGATTACAAGGCAATG 1140
I K P I L G K Y Y P F D G T P I Y K A M
TGGAGGGAGGCAAAAGAGTGCCTTTACGTTGAGCCTGACGTTGGTGGTGGTGGTGGT 1200
W R E A K E C L Y V E P D V G G G G G G
AGCAAAGGTGTTTTTTTGGTATCGTAACAAGTTCTAAAGACCGACCAACTGCCTGATAGCT 1260
S K G V F W Y R N K F *
GGCCGGCGAAATCAACGTAAACGTACTTATTAGACTAGTGTTAACTAGGGAAGTTAATA 1320
ATTAATGGTAGGAAAATGTGGAATAGTTGCCTAGTAGTTTTATGTATTAAGTGTGTATT 1380
AATAAACTATATGTTAGTAAAAA 1411

```

Fig. 5.1 Nucleotide and deduced amino acid sequences of *pghD12-1*. The translation start site (ATG) is in bold type. The oligonucleotide, *d12A4*, used as PCR primer for the amplification of 5' ends is underlined and its 5' to 3' direction is indicated by an arrow pointing to 3' direction. The stop codon is indicated by an asterisk.

5.3.1.3 Amplification of 5' upstream sequence of *pghD12-2* from genomic DNA

The PCR approach using λ cDNA library DNA extracts as template was also employed to obtain the missing 5' end of *pghD12-2*, but it proved to be unsuccessful. Instead, by using a primer-adaptor mediated PCR method, with the combination of BglIII adaptor-primer and a cDNA-specific primer, *3111a1*, a DNA fragment containing the expected 195 bp missing at the 5' end of *pghD12-2* and an additional 621 bp upstream sequence was successfully amplified from the genomic DNA of *G. hirsutum* cv Deltapine-16. Fig. 5.2 illustrates the complete sequence of *pghD12-2* which contains an ORF of 384 amino acid residues, 621 bp 5' upstream sequence and 173 bp of 3' UTR.

A putative TATA box (TATAAA) was located at 544 bp upstream from the translational start site. It is estimated that the transcription usually starts at about 30 bp downstream the TATA box (Zhu, 1996). However, it was not clear whether the 5' UTR of *pghD12-2* also contains an intron as it was revealed in *pghD12-1* (it will be discussed in the following chapter). In order to clarify the identity of this region, a PCR strategy was employed using the crude λ DNA extract from a cottonseed cDNA library as DNA template and a pair of PCR primers based on the known sequences from this region. A sense primer, *D125-1* (5'- TTTCTGTTTAATT-GGGTACC -3') starting from 54 bp downstream the putative TATA box, in combination with an antisense primer *3111a1* was used in the PCR. A product with exactly the same sequence identity with *pghD12-2* in this region was obtained. This indicated *pghD12-2* does not contain an 5'UTR intron and it has an approximately 500 bp 5'UTR, although the exact transcription start site was not determined.

AGCACTCTCCAGCCTCTCACCGCAGATCTTTATCTACCTTAGTGTTTTAATTGAAATAGC -562
TAGTAGTGTTAACTATTTGGAATAACTAATGATAT**TATA**AAAAAAGGTAAATATATTTTC -502
AATGTTCAAAATAGTACCTAAGATTTTTT**TTTCTGTTTAATTGGGTACCTAAGCTTGTCCT** -442
D125-1 primer
CAAGGTTCAATTTGGTACCTAAATTTTGCTTTTAGTTTAAATTAGTACTTGAACCTTGGC -382
TTGATAGGTCAAATTTGGTCAAATAAGTCTTTCTTTTGTAGAAATCAAATAAGTCATTTATT -322
GTACATACTAATTTGGACTAAAAACCAAATTCATGTACTAATTCAAACCTAAGAAATCAAG -262
TTCAGATGCCTAATTTGACTAAAAATAAATTTTAGGTACTAAATTAACCTTTAAAGTCAAG -202
TTCAAGTACCTAATTTGAACAAAAAAATCTGAAGTACTAATTTAAACCTGATAGTCAAGT -142
TCAAGTATCAAATGTATATTTATCAACTAATATGGTGTGTGTAGGGGTTTAGAGGTAAAG -82
ATTGAGTAATTAACATGGAGAAACCATGTTTGTCTTTTAACCATTTCCACTATCCATTT -22
ATTTTTTTCAGGGTGTGGAACA**ATCGGTGCTGGAGGCAGAATGTCGGTTCCAACGAGTCCA** 39
M G A G G R M S V P T S P
AAAAAACCCGAATTCAACTCACTGAAGCGAGTTCCATACTCAAAGCCACCCTTCACTCTG 99
K K P E F N S L K R V P Y S K P P F T L
AGTGAAATCAAGAAAGCCATCCCAACCACTGTTTCCAGCGCTCCGTTTTACGCTCATTC 159
S E I K K A I P P H C F Q R S V L R S F
TCATATCTCCTTTACGACTTTATATTTGGCTCTCTTTTTTTACCATGTGGCCACCAATTAC 219
S Y L L Y D F I L A S L F Y H V A T N Y
TTCCCTAACCTTCTCAGGCTCTCTCCAACGTGGCTTGGCTCTTTATTGGGCCATGCAA 279
F P N L P Q A L S N V A W P L Y W A M Q
GGTTGCATTTTGACCGCGCTTGGGTCATAGCCCATGAATGTGGCCACCATGCTTTTCAGT 339
3111a1 primer
G C I L T G V W V I A H E C G H H A F S
GATTATCAATGGCTTGACGACACCGTGCGCTTATCCTCCACTCTTCTCTTAGTTCCA 399
D Y Q W L D D T V G L I L H S S L L V P
TATTTCTCTTGAAATATAGCCACCGCGTCACCATTCTAACACCGGTTCCCTCGAAAGG 459
Y F S W T K Y S H R R H H S N T G S L E R
GATGAAGTGTTTCGTTCCCAAGAAAAATCTGGTTTAAGATGGTGGGCCAAACACTCAAC 519
D E V F V P K K K S G L R W W A K H F N
AATCCACCGGGTCGGTTTTCTGTCAATCACCATTCAACTTACCCTTGGTTGGCCGCTTTAC 579
N P P G R F L S I T I Q L T L G W P L Y
TTAGCTTTCAACGTTGCCGCGCGCTTACGACAGGTTTCGTTGCCACTATGACCCTTAC 639
L A F N V A G R P Y D R F A C H Y D P Y
GGCCCCATATTTTCCGACCGGAACGACTCCAAATCTATATCTCTGACGCCGCGCTCCTC 699
G P I F S D R E R L Q I Y I S D A G V L
GCTGTGCGCTATGCGCTCTACCGTCTCGTGTTGGCCAAAGGGGTAGGTTGGGTTATTAGC 759
A V A Y A L Y R L V L A K G V G W V I S
GTTTATGGGGTGCCATTATTGGTGGTTAACGCCTTCTTAGTAATGATCACGTATTTGCAA 819
V Y G V P L L V V N A F L V M I T Y L Q
CACACTCACCCATCTTTGCCGCACTATGATTCCTCGGAGTGGGACTGGATGAGAGGAGCT 879
H T H P S L P H Y D S S E W D W M R G A
TTATCAACTGTGGACAGAGATTATGGGATTTTAAACAAGGTTTTCCATAACATAACCGAC 939
L S T V D R D Y G I L N K V F H N I T D
ACTCATGTGGCTCATCATTTGTTTTTCGACAATGCCTCACTATCATGCCATGGTGGCCACC 999
T H V A H H L F S T M P H Y H A M V A T
AAGGCGATAAAGCCCATATTGGGGGAATACTATCAGTTCGATGGGATGCCTGTCTATAAG 1059
K A I K P I L G E Y Y Q F D G M P V Y K
GCGATATGGAGGGAGGCGAAGGAGTGTCCTACGTTGAACCAGATGAGGGCGACAAGGAT 1119
A I W R E A K E C L Y V E P D E G D K D
AAAGGTGTGTTTTGGTTTAGAAACAAGCTTTAAATATTTGCATTTTACCTTAGGCATGTT 1179
K G V F W F R N K L *
CTAGTCGTTGATGTTTTAAGGATATTTTAGCCGACATACTTGGTTTTCTTTTTGGGACT 1239
TTTTAGCTTTGTATTTGCAGACAATAATCTTGTTCATTTAAATAATGGTAGAAATAAA 1299
TACACAGCATGGATTGGCAATAAAAA 1325

Fig. 5.2 Nucleotide and deduced amino acid sequences of pghD12-2. The putative TATA box and translation start site are in bold type. The *D125-1* and *3111a1* primers are underlined. The stop codon is indicated by an asterisk.

5.3.2 Evaluation of Gene Expression by RNA Northern Blot Analysis

Transcript levels of the pghD12-1 and pghD12-2 were evaluated in developing embryos and leaf tissues in *G. hirsutum* by RNA Northern blot analysis. As shown in Fig. 5.3B, the transcription of pghD12-1 was already induced in the earliest sampled embryos (25 DAA). The expression level rapidly increased as embryo development advanced and the highest expression being at 30 DAA. Significantly lower amounts of transcripts were detected in the embryos near maturity (45 DAA). No transcripts of the pghD12-1 were detectable in cotton leaf tissues. In contrast, very low expression of pghD12-2 was detected in both developing embryos at all stages and leaf tissues (Fig. 5.3C).

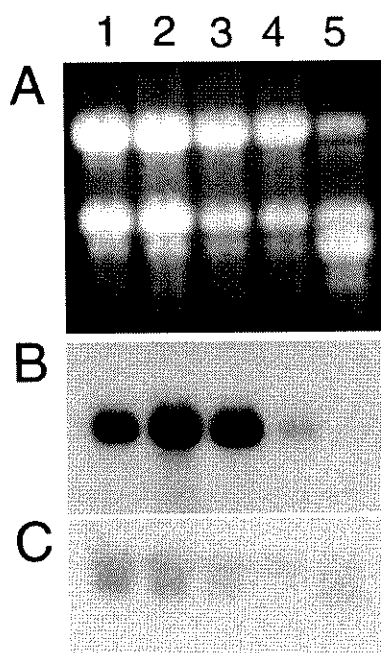


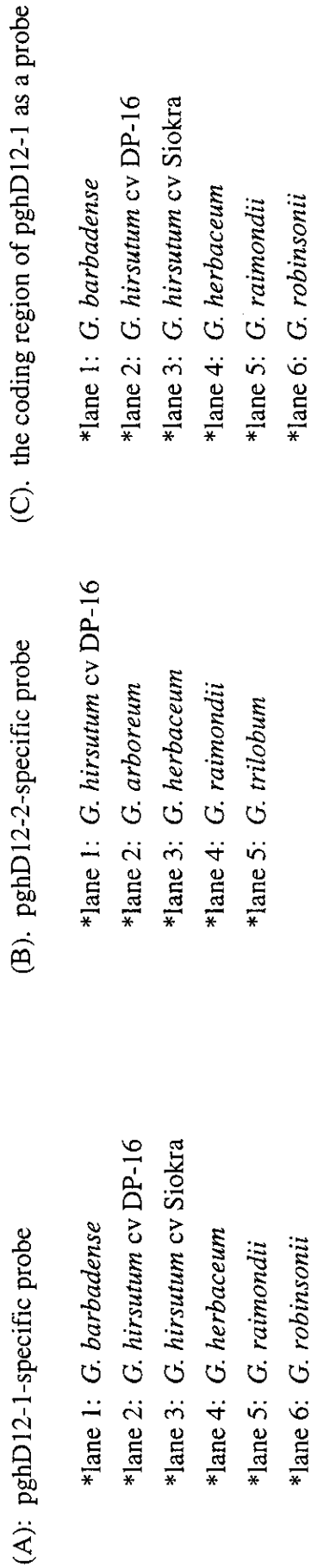
Fig. 5.3 Northern blot analysis of developing embryos and leaf tissues from *G. hirsutum* cv Deltapine-16. (A) EtBr-stained formaldehyde denaturing RNA gel. Lanes 1-4 represent RNA samples isolated from embryos of 25, 30, 36, 45 DAA, respectively. Lane 5 represents RNA isolated from young leaves. (B). RNA gel shown in panel A was transferred to a Hybond N⁺ membrane and hybridised by an [α -³²P]dCTP-labelled pghD12-1-specific probe (3' UTR of pghD12-1). (C). RNA gel shown in panel A was transferred to a Hybond N⁺ membrane and hybridised by an [α -³²P]dCTP-labelled pghD12-2-specific probe (3' UTR of pghD12-2).

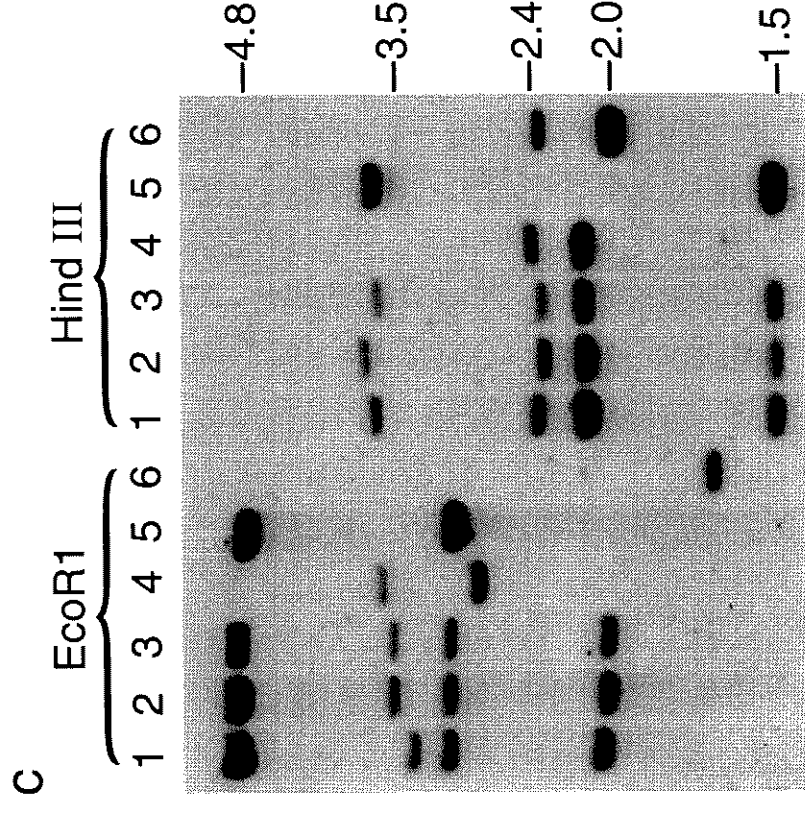
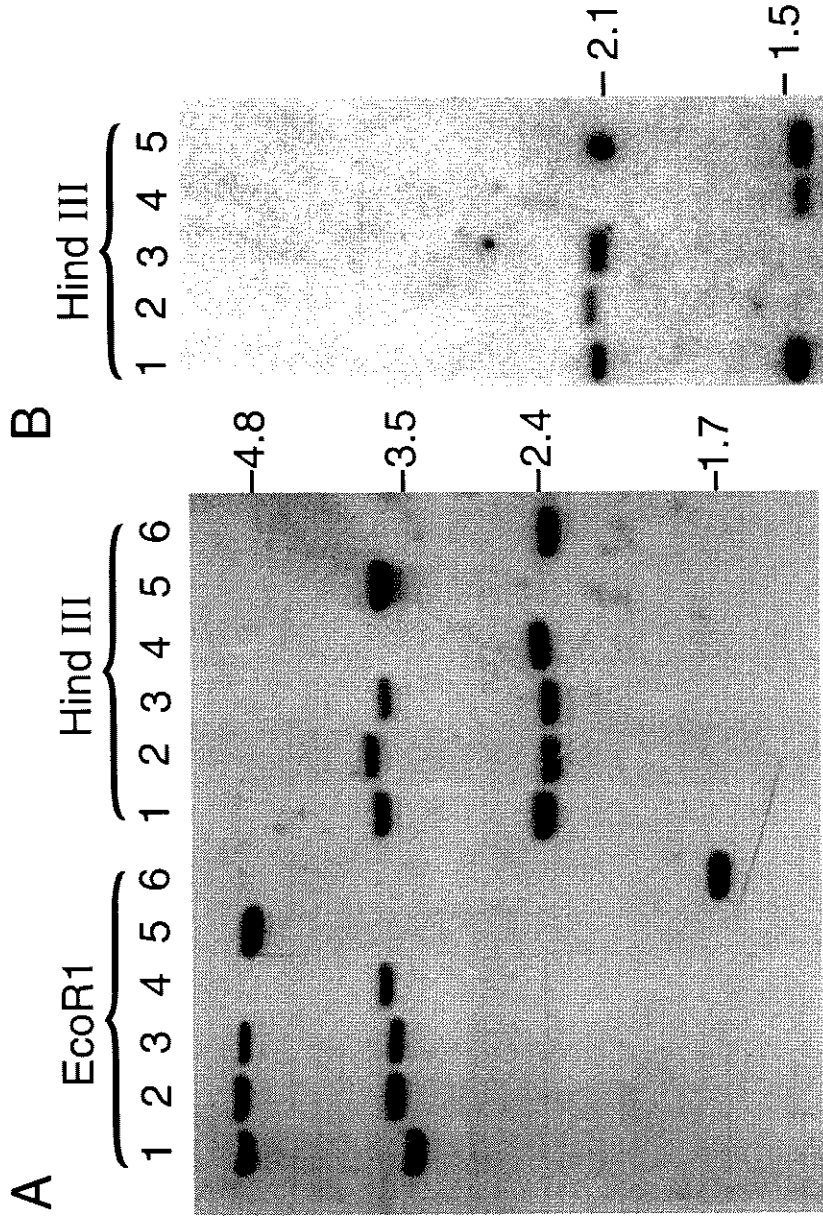
5.3.3 Genomic Southern Blot Analysis in *Gossypium* Species

Fig. 5.4A represents a Southern blot probed with a pghD12-1-specific fragment based on the sequence of the 3' UTR of pghD12-1. In both EcoRI and HindIII digests, there are two distinct bands in tetraploids (lane 1-3) and one band in each of the diploids (lane 4-6). Of the two bands in tetraploids, the lower band is similar in size to that of *G. herbaceum* (lane 4) and the upper band similar to that of *G. raimondii* (lane 5). A single band is also present in *G. robinsonii*, the C-genome diploid species. Fig. 5.4B illustrates the pghD12-2-specific/HindIII pattern of a tetraploid species, *G. hirsutum* (lane 1), two A-genome diploids (lane 2-3) and two D-genome diploids (lane 4-5). Similar to the pghD12-1 patterns, there are two pghD12-2 bands in *G. hirsutum* and a single band in each of the diploids. Of the two bands in *G. hirsutum*, the upper band is similar to the A-genome bands and the lower band is similar to the D-genome bands.

Fig. 5.4C represents the Southern blot analysis using the more conserved coding region of pghD12-1 as a probe. Five (EcoRI digests) and four (HindIII digests) distinct bands can be distinguished in the tetraploid cottons (lanes 1-3) while only two bands are discernible in the diploids (lanes 4-6). However, one HindIII band (the third band from top in lanes 1-4; the second band in lane 6) appears to be at least twice as strong as the other bands. Hence, it is likely that this band contains multiple copies of the gene. Therefore, it can be suggested that there are at least five members of the microsomal ω -6 desaturase genes in each of the tetraploid cotton species. Further, it would appear that A-genome diploid species *G. herbaceum* (lane 4) has at least three and the D-genome diploid species *G. raimondii* (lane 5) has two members of the gene family. The inconsistency between the banding patterns obtained with the EcoRI and HindIII-digested *G. robinsonii* DNA (lane 6) is perhaps due to the incomplete digestion of the *G. robinsonii* genomic DNA by EcoRI. It is also noteworthy that the HindIII pattern observed in *G. robinsonii* shows high degree of similarity to that of the A-genome diploid species.

Fig. 5.4 Southern blots of cotton genomic DNAs probed with cotton microsomal ω -6 desaturase cDNA fragments.





5.4 DISCUSSION

5.4.1 Sequence Features of cDNAs for Microsomal ω -6 Desaturases in *G. hirsutum*

In this Chapter, the isolation and characterisation of two different cDNAs, pghD12-1 and pghD12-2 encoding microsomal ω -6 desaturase in *G. hirsutum* were described. The sequence comparison between pghD12-1 and pghD12-2 revealed that they possess unique 5' and 3' UTRs, indicating that they have evolved independently since their divergence from a common ancestral gene. The gene specific probes based on their unique 3' UTR sequences have been used for the Southern blot and Northern blot analyses in this study.

A feature of the nucleotide sequence of the two full length cDNAs, pghD12-1 and pghD12-2 is the differences in the regions surrounding the translation initiation codon ATG, which is illustrated in Fig. 5.5.

pghD12-1	CAAA <u>AT</u> GGG
pghD12-2	AACA <u>AT</u> GGG
plant consensus	AACA <u>AT</u> GGC

Fig. 5.5 Alignment of nucleotides surrounding the translation initiation codon ATG of the two full length cDNAs encoding for cotton microsomal ω -6 desaturases and the plant consensus motif for translation initiation (Lütcke et al., 1987). The translation initiation codon is underlined.

Translation usually begins at an ATG codon, and the efficiency of ATG codon recognition is modulated by the sequence context of this codon (Kozak, 1986). Sequence compilation of plant translation initiation regions has suggested that the most frequent (and therefore regarded as optimal) ATG context for plant genes is similar to that of mammalian genes (AACAAATGGC) (Cavener and Ray, 1991; Lütcke et al., 1987). The most crucial nucleotides in both cases are a purine at position -3 and a guanine at position +4 (where the A of ATG is +1). Start codons that deviate from the optimal context at one or more of the crucial positions may be recognised less efficiently and allow the passage of ribosomes to start codons further downstream (leaky scanning). The influence of sequence context on initiation in plant cells has been directly verified by mutagenesis studies *in vitro*, in plant protoplasts and in transgenic cells (Dinesh-Kumar and Miller, 1993; Kozak, 1989; Murray et al., 1991). The translation initiation codon of pghD12-2 appears to be optimal, eight out of nine base pairs of the initiation codon are identical with the plant consensus. In pghD12-1, seven out of nine bases are identical to the plant consensus, but one of the most crucial bases, the purine at position -3 of consensus has been replaced by an adenosine. However, as it will be discussed below, the pghD12-1 is a highly expressed gene during cottonseed embryo development concurrent with the oil biosynthesis in cottonseeds, whereas the pghD12-2 is poorly expressed in both embryo and leaf tissues. The optimal codon context in the low expression gene, pghD12-2 and sub-optimal context in the highly expressed pghD12-1 contradicts the previous findings about the role of the translation initiation codon context in the level of gene expression. A similar observation has been made in soybean lipoxygenase gene (Kato et al., 1991). Its sub-optimal codon context ORFs (AAAGATGTTT) was found to give 10 times higher GUS expression rates than the context in the standard GUS expression plasmid pBI121. Therefore, the context of translation initiation codon has probably only minor influences on the translation efficiency of the cotton microsomal ω -6 desaturase genes.

5.4.2 Comparison of the Deduced Polypeptide Sequences of *G. hirsutum* ω -6 Desaturase and Other Plant Membrane-bound Fatty Acid Desaturases

As shown in Fig. 5.6, the alignment of the microsomal ω -6 desaturases from cotton (pghD12-1 and pghD12-2), soybean (gmFAD2-1, FAD2-2), *A. thaliana* (atFAD2), microsomal ω -3 desaturase from *B. napus* (bnFAD3) and plastidial ω -6 desaturase from soybean (gmFAD6) shows the highest sequence identity among the five microsomal ω -6 desaturases in general. When inspected closely, the two seed-specific genes, cotton pghD12-1 and soybean gmFAD2-1 share more sequence identity than the others, while cotton pghD12-2 is most similar to the constitutively expressed soybean gmFAD2-2. Significantly less sequence identity is revealed between the *B. napus* microsomal ω -3 desaturase bnFAD3 and all of the microsomal ω -6 desaturases. The plastidial ω -6 desaturase, gmFAD6, remains the least similar sequence to all the microsomal desaturase amino acid sequences. For instance, pghD12-2 shared 74, 76, 71, 77% amino acid identity with pghD12-1, atFAD2, gmFAD2-1, gmFAD2-2, respectively; but only 40 and 27% identity with bnFAD3, gmFAD6, respectively. Previous studies confirmed that microsomal and plastidial types of ω -3 desaturase share reasonably high sequence identity and both types are also similar to the microsomal ω -6 desaturase genes in plants (Yadav et al., 1993). However, the plastidial ω -6 desaturase is very different from its microsomal counterpart (Falcone et al., 1994). This may indicate that the plastidial ω -6 desaturase evolved independently long before the divergence of microsomal ω -6 desaturase and both types of ω -3 desaturases.

The three histidine boxes which were shown to be highly conserved in all membrane-bound desaturases (Schmidt et al., 1994) were also observed in the *G. hirsutum* microsomal ω -6 desaturases. As shown in Fig. 5.6, among the five microsomal ω -6 desaturases listed, the deduced amino acid residues for the three boxes are HECGHH (with exception that pghD12-1 has HEWGHH), HRRHH and HVAHH, respectively.

Although the sequence consensus fell away in comparison with other bnFAD3 and gmFAD6, the general sequence HXXXH remains conserved.

The distance between the first and second histidine boxes is 31 amino acid residues among all protein sequence listed in Fig. 5.6. The distance between the second and the third boxes is also conserved with 169 amino acid residues among the five microsomal ω -6 desaturases, but 162 and 154 amino acid residues for the bnFAD3 and gmFAD6, respectively. The conservation of distances between histidine boxes may emphasise their importance in three dimensional structure of microsomal ω -6 desaturase enzymes and their close cooperation in the catalysing process.

Similar to other microsomal ω -6 desaturases, the deduced protein sequences of pghD12-1 and pghD12-2 lack a recognisable signal sequence for targeting to the ER. It is possible that the gene products may insert into ER post-translationally as does the rat liver Δ^9 stearoyl-CoA desaturase (Thiede et al., 1985). Furthermore, a lysine-rich carboxyl-terminal motif present in *A. thaliana* FAD3 enzyme has been suggested to represent the retention signal of integral membrane proteins in ER (Jackson et al., 1990). But it is found in neither cotton nor other plant microsomal ω -6 fatty acid desaturases. However, near the C-terminal of the pghD12-1 ORF there is a stretch of six continuous glycine residues which makes it distinguishable from other microsomal ω -6 desaturases (underlined in Fig. 5.6). The significance of such an unique sequence feature of pghD12-1 remains to be investigated.

Fig. 5.6 Comparison of the deduced amino acid sequences of five microsomal ω -6 desaturases and other membrane-bound fatty acid desaturases, including microsomal ω -3 desaturase and plastidial ω -6 desaturase.

gmFAD2-2: *G. max* microsomal ω -6 desaturase
pghD12-2: *G. hirsutum* microsomal ω -6 desaturase
atFAD2: *A. thaliana* microsomal ω -6 desaturase
gmFAD2-1: *G. max* microsomal ω -6 desaturase
pghD12-1: *G. hirsutum* microsomal ω -6 desaturase
bnFAD3: *B. napus* microsomal ω -3 desaturase
gmFAD6: *G. max* plastidial ω -6 desaturase

*Shaded areas indicate homologous residues in one or more of the different desaturase sequences.

*Bold face residues indicate amino acids conserved among all sequences.

*Dots indicate spaces introduced to maximise alignments.

*The three histidine boxes proposed to be important in iron binding are underlined.

*The six glycine repeats in relative C-terminus of pghD12-1 enzyme is also underlined.

gmFAD2-2 1 MGAGGRTDVPPAN KSEVDF RVP
pghd12-2 1 MGAGGRMSVPTSP KSEFNN RVP
atFAD2 1 MGAGGRMPVPTSS KSETD T RVP
gmFAD2-1 1 MGLAKETMGGGRVAKV ZVQKKF SRVP
pghd12-1 1 MGAGGRMPDGI ZEMKGF SRVP
bnFAD3 1 MGVAMDQSNANGDE FDF
gmFAD6 1 MACTLADSLLFKGSYQKPVLRDRDIAARYSPGIFSLNSGLIQKRFRQRNFVTRNKVTIHAIAIPVQPAPEAEYRK

26 FEKQPF SQIKKAIPPHCFQRS LRFSYV D T AFCLYYVATMYFEL LPQPLS RQMA YWA QGC LTGV
pghd12-2 26 YSKPPF SIKKAIPPHCFQRS LRFSYV D I ASLPYHVATMYFN LPQALSNAWPLYWAQGC LTGV
atFAD2 26 CEKPPF GDKKAIPPHCFQRS PRFSYV SD I ASCLYYVATMYFS LPQPLS AWPPLYWACQGC LTG
gmFAD2-1 31 NTKPPF GQKKAIPPHCFQRS LRFSYV D S AFIYYATMYFEL LPQPSL AWPYMW QGC LTGV
pghd12-1 25 IEKPPF GOIKQAIPPHCFQRS LRFSYVHD C ASFFYYATMYFEL LPQPS AWPYMW QGC LTGV
bnFAD3 20 SAQPPF KGDIAAIPPHCFV SPLRSM SVARD FAVVALAVAVED S FFWPLYWAAQGT FM
gmFAD6 81 QLAED GFRQ GEP SDDVTLD NPLPREV DDVKAWSVLISVTYALCF ISKAP YLLPLM VNTGTA TGP

101 WVIAHECGHHAFSDYQL LDD VGLILHS ALLVPYFSWK SHRRHHSNTGSLERDEVFPKQKSC W SKYLNPNP GRLL
pghd12-2 101 WVIAHECGHHAFSDYQWLDD TVGLILHS LLVPYFSWK SHRRHHSNTGSLERDEVFPKQKSA W AKHFNPNP GRLL
atFAD2 101 WVIAHECGHHAFSDYQWLDD TVGLILHS LLVPYFSWK SHRRHHSNTGSLERDEVFPKQKSA W SKYLNPNP GRLL
gmFAD2-1 105 WVIAHECGHHAFSDYQW DDVGLILHS LLVPYFSWK SHRRHHSNTGSLERDEVFPKQKSA W SKYLNPNP GRLL
pghd12-1 100 WVIAHECGHHAFSDYQW DDVGLILHS LLVPYFSWK SHRRHHSNTGSLERDEVFPKQKSA W SKYLNPNP GRLL
bnFAD3 88 VICHG CGHG FSDIPL LTA VGHILHS LLVPYHGW ISHRTHHQNHGK ENDES VLPKLY NLS HSTRLL
gmFAD6 161 VICHG CH FSNKL DIVGT AFM FL YP YEPW NHDRHH KTNML EETA HPVKNDEFES TPL

181 LGWPLYLALNVSGRPYDRFA CHYDPY PIYSDRERLOIYISDAG LAVYGLRLMAK GLW CVYGV
pghd12-2 181 LGWPLYLALNVSGRPYDRFA CHYDPY PIYSDRERLOIYISDAG LAVYGLRLMAK GLW CVYGV
atFAD2 181 LGWPLYLALNVSGRPYDRFA CHYDPY PIYSDRERLOIYISDAG LAVYGLRLMAK GLW CVYGV
gmFAD2-1 185 LGWPLYLALNVSGRPYDRFA CHYDPY PIYSDRERLOIYISDAG LAVYGLRLMAK GLW CVYGV
pghd12-1 180 LGWPLYLALNVSGRPYDRFA CHYDPY PIYSDRERLOIYISDAG LAVYGLRLMAK GLW CVYGV
bnFAD3 163 RYTVPLPMLA PLYLW RSPGE SHYNPYSE PSEKLIATSTCWSI ATLYLGLV GPVTYKVYGV
gmFAD6 232 KKA IYGYOP RCN IY HWMWHDL KKRPSSEVPRAKISLAC TAF AIG PLIYKTG MGV KPLLP

259 LLVN FLV ITLQH THPALPHYTSSEWDLRGAL TVDRDYGILNKKVFHNI TDTHVAHHLFS T MPHYHAMEATKA
pghd12-2 259 LLVN FLV ITLQH THPSLPHYDSSEWDLRGAL TVDRDYGILNKKVFHNI TDTHVAHHLFS T MPHYHAMEATKA
atFAD2 259 LLVN FLV ITLQH THPSLPHYDSSEWDLRGAL TVDRDYGILNKKVFHNI TDTHVAHHLFS T MPHYHAMEATKA
gmFAD2-1 263 LLVN FLV ITLQH THPALPHYDSSEWDLRGAL TVDRDYGILNKKVFHNI TDTHVAHHLFS T MPHYHAMEATKA
pghd12-1 258 LLVN FLV ITLQH THPALPHYDSSEWDLRGAL TVDRDYGILNKKVFHNI TDTHVAHHLFS T MPHYHAMEATKA
bnFAD3 238 YIFVM FLDAVTLHHH GHDDKLPYRGKEWSLRGCL TDRDYGIFNN EHCI GTHVHHLEFPQPHYHL DATK
gmFAD6 304 WGYHT STFTMHH TAPYPPKYSEWRAQQLNGTVHCDYPKWIE LCHDI VHPHH SPRPSYNNLRAAHEK

336 IKPILGY YRFDTPPVKA WREA ECLYVEPD QSTESKGV WYNNKL*
pghd12-2 336 IKPILGY YQDGMPPVKA WREA ECLYVEPD EGDDKGV WNNKL*
atFAD2 336 IKPILGY YQFDGTPWYVA WREA ECLYVEPD REGDKGV WYNNKL*
gmFAD2-1 340 IKPILGY YQFDDTPFYKA WREA ECLYVEPD EGTSKGV WYNNKY*
pghd12-1 335 IKPILGY YQDGTPIYKA WREA ECLYVEPD VGGGGGSGKV WYNNKY*
bnFAD3 316 AKENLGRY YREPKTSGAIF HLVESLVAS KNDHYSDTDIVF ETDPDLYVYASDKSKIN*
gmFAD6 382 QENWGQY LNEASWN RLMKTI MTVCQVYKEKSLCCLNATCP*
.....

5.4.3 Hydropathy Analysis of pghD12-1 and pghD12-2

The putative cotton microsomal ω -6 desaturases (pghD12-1 and pghD12-2) possess six transmembrane helices (numbered from 1 through to 6, in Fig. 5.7). This fits perfectly the model proposed by Stuke et al. (1990) by comparing the yeast and rat liver Δ^9 stearoyl-CoA fatty acid desaturases. According to that model, the first two (1-2) and last two (5-6) helices form two membrane-traversing loops, each consisting of two transmembrane helices, whereas the remaining two helices (3-4) are single-pass membrane spanning segments. A similar observation has also been made for the microsomal ω -6 desaturase in *A. thaliana*, FAD2 (Okuley et al., 1994). Interestingly, the three histidine boxes proposed to be iron-binding sites during catalysis are not located in the region of transmembrane helices. As indicated in Fig. 5.7, the first two boxes are located between helices 2 and 3, 3 and 4, respectively. The third box is at the C-terminal part of the protein past the second membrane-traversing loop. In fact, the areas surrounding these three boxes have intermediate polarity and as such are optimal as reaction centres which have to interact with hydrophilic electron donors as well as with the hydrophobic fatty acid substrates to be desaturated. The positions of these boxes become important since it determines whether or not the boxes will be placed in the same side of membrane and is therefore crucial for the understanding of its catalysing mechanism regarding electron donor docking, electron transfer etc. According to Okuley et al. (1994), the second histidine box of FAD2 will always be at the side of membrane away from the remaining two boxes. This is probably also the case for the *G. hirsutum* microsomal ω -6 desaturases. However, it needs to be confirmed by further studies using other research methodologies, such as direct measurement of membrane topology (Jennings, 1989).

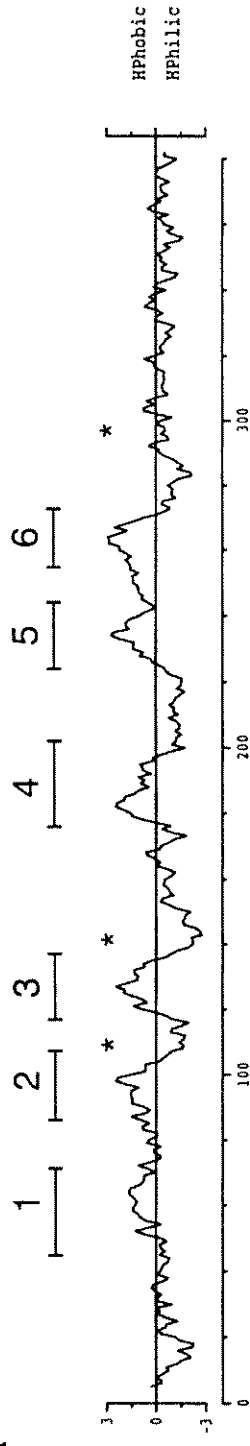
Fig. 5.7 Hydropathy analysis of the microsomal ω -6 desaturase in *G. hirsutum* using Kyte-Doolittle method.

(A). pghD12-1
(B). pghD12-2

*The presumptive six transmembrane helices are indicated by bold lines above those regions.

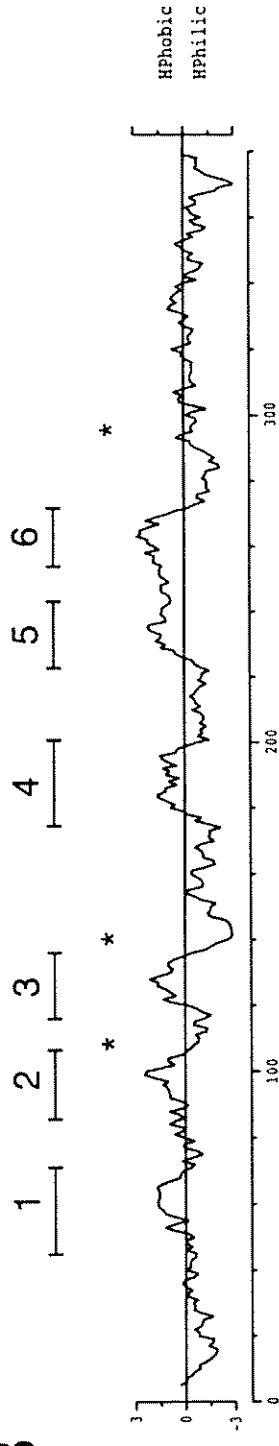
*The three iron-binding histidine boxes are indicated by asterisks.

A



Amino acid residue number

B



Amino acid residue number

5.4.4 PCR Strategies for Rapid Amplification of 5' cDNA Ends

Two different PCR strategies were successfully used for amplifying the 5' ends which were missing in the cDNAs isolated from cDNA library. Since pghD12-1 is a seed-specific gene and it is highly expressed in embryo tissues, a simple PCR was carried out using crude DNA extracts from λ ZAPII cDNA library without phagemid excision as DNA template, and a gene specific primer in combination with pBluescript vector-specific primers. This protocol allowed successful amplification of 5' end of pghD12-1 in a single step. However, the amplification of 5' end of pghD12-2 proved to be difficult. The above method failed even with nested primers. This is perhaps because pghD12-2 is a low expression gene and poorly represented in the cottonseed cDNA library. Other PCR approaches using genomic DNA as template, including inverse PCR (Ochman et al., 1988) and random primed gene walker PCR (Trueba and Johnson, 1996) also failed to yield the missing 5' portion of pghD12-2. The targeted DNA sequence was finally obtained using a simplified procedure with the principle of primer-adapter mediated PCR (Espelund and Jakobsen, 1992). In the original protocol of primer-adapter mediated PCR, the target sequence was amplified with biotinylated primers and the PCR product was separated with the aid of streptavidin coated magnetic beads. This is not only expensive and technically demanding, but also time consuming. By carefully adjusting the ratio of restricted genomic DNA and the adapter, coupling with a hot start PCR procedure, it is possible to omit the use of biotin/streptavidin. By choosing the correct restriction enzyme which cut the DNA not too far from the gene-specific primer, or using long range PCR techniques, the simplified adapter-primer mediated PCR will prove to be an efficient protocol for amplifying DNA fragments from low-abundance genes.

5.4.5 Gene Expression Patterns of Cotton Microsomal ω -6 Desaturase

Microsomal ω -6 desaturase is encoded by a single gene in *Arabidopsis*, FAD2, and this gene is expressed in both vegetative tissues and developing embryos and is presumably responsible for conversion of 18:1 to 18:2 fatty acids in both membrane and seed storage lipids (Okuley et al., 1994). Previous studies with the *A. thaliana fad2* mutants deficient in activity of the microsomal ω -6 desaturase have shown that the growth of vegetative tissues was severely affected at low temperature (Miquel et al., 1993). Further studies also indicated that *fad2* seed failed to develop or germinate at low temperature (Miquel et al., 1994). It is probable that a high level of polyunsaturated fatty acid may be required for maintaining membrane fluidity and membrane-bound enzyme activity for cell division and differentiation during early seed development. Therefore, genetic engineering of plant seed oil using such a constitutively expressed gene clone may result in plant lines with unacceptable performance in the field.

A small gene family of two distinct members has been reported in soybean. Similar to *A. thaliana* FAD2, the soybean FAD2-2 was constitutively expressed in leaf and embryo tissues (Heppard et al., 1996), while FAD2-1 is specifically induced during embryo development concomitant to seed storage lipid synthesis. FAD2-1 was successfully used to suppress its endogenous counterpart and achieved the high-oleate soybean oil without detrimental effect on soybean vegetative growth (Kinney, 1995).

Similar to the scenario in soybean, Northern blot analysis of cotton microsomal ω -6 fatty acid desaturase genes in this study revealed that the pghD12-1 gene was not expressed in leaf tissue, but specifically activated in developing cotton embryos. It is likely, therefore, pghD12-1 is responsible for the introduction of the second double bond into the fatty acid chain in cottonseed oil. In contrast, pghD12-2, has very low expression in both cotton leaf tissues and developing embryos. Therefore, it may not play a major role in the desaturation of cottonseed storage lipids and membrane lipids.

Its cellular localisation, defined expression pattern and functional aspects is yet to be determined.

The seed storage proteins in cottonseed are generally believed to accumulate earlier than storage lipids (Elmore and Leffler, 1976; El-Nockrashy et al., 1976; King and Leffler, 1979). The storage lipids increase rapidly in cottonseed at about 25-40 DAA and then increase slowly, at a more or less continuous rate associated with the weight increase of the embryo (Chapter Three in this thesis; Brown and Kurtz, 1959; El-Nockrashy et al., 1976; Touma-Touchan, 1977). The Northern blot analysis of pghD12-1 gene in this study indicated that the expression of pghD12-1 has already been highly induced at 25 DAA, and highest expression was observed at 30-36 DAA, the transcripts of pghD12-1 was drastically reduced at 45 DAA. The expression level of pghD12-1 significantly increased during the period of maximum storage lipid synthesis and accumulation, which is consistent with the increased linoleic acid accumulation during the same period (Chapter Three).

5.4.6 Estimation of Gene Copy Number and the Size of Microsomal ω -6 Desaturase Multigene Family

Southern blot analyses indicate that there are at least five closely related genes in each of two tetraploids, *G. barbadense* and *G. hirsutum*. The conservation of polymorphic restriction fragment patterns in tetraploid genomes and those representing their putative progenitors was sufficient to allow the assignment of each of the fragments in allotetraploid to their respective subgenomes. It is estimated that in tetraploid cotton, at least three genes (copies) were from the putative A-genome progenitor species and at least two may be contributed by the D-genome progenitor. Southern blot analysis also indicated that pghD12-1 and pghD12-2 are nonallelic and there are at least two copies of each of the genes in the tetraploid cotton species and at least one copy in diploid species.

Southern blots of soybean genomic DNA probed with FAD2-1- and FAD2-2-specific sequence fragments indicated that there are at least two copies of each of these genes in soybean genome (Heppard et al., 1996). However, no evidence was presented to estimate the size of the gene family. More recently, microsomal ω -6 desaturase gene family in *B. napus* has been estimated to contain 4-6 gene copies per haploid genome and they originated from gene duplications or triplications in the progenitor species prior to the formation of *B. napus* (Scheffler et al., 1997). In the current study in cotton, both pghD12-1 and pghD12-2 appear to be single-copy genes in diploid cotton species, indicating the gene duplication observed in *B. napus* did not occur in cotton. This is consistent with the observation made in the Δ 9 stearoyl-ACP desaturase gene in cotton, although the size of Δ 9 stearoyl-ACP gene family is much larger (Chapter Four).

Chapter Six

Evolutionary Studies of pghD12-1 Gene in *Gossypium* species on the basis of its 5'UTR Intron DNA Sequence

6.1 INTRODUCTION

The genus *Gossypium* has approximately 50 species of shrubs and small trees distributed throughout the tropics and subtropics (Fryxell, 1992). Global radiation of the genus *Gossypium* was accompanied by extensive chromosomal evolution and impressive diversification in morphology and ecology. The collective observations of chromosome size, pairing behaviour, and relative fertility in interspecific hybrids have led to the recognition of eight diploid genome groups designated A through G, plus K (Endrizzi et al., 1985; Stewart, 1995). The New World tetraploids (AADD) are formed by the merger of A- and D-genomes as a result of interspecific hybridisation. Although the actual progenitor diploid taxa are most likely extinct, the A-genome donor is similar to *G. herbaceum* and the best model of the ancestral D-genome parent is *G. raimondii* (Endrizzi et al., 1985). Following polyploidisation, the newly formed allopolyploids diverged into five modern species (De Jode and Wendel, 1992; Brubaker and Wendel 1993; Wendel et al., 1994), including the commercially important *G. hirsutum* and *G. barbadense*.

The recent development of molecular methods has greatly increased the data available for phylogenetic inference. Phylogenetic investigations in *Gossypium* used several different molecular data sets, including cpDNA restriction sites (Wendel and Albert, 1992), DNA sequence for the 5S ribosomal genes and spacers (Cronn et al., 1996), DNA sequence data for the cpDNA gene *ndhF* and the nuclear 5.8S gene and its flanking internal transcribed spacers (Seelanan et al., 1997). In these analyses the genealogical lineages of species are largely congruent with geographical distributions and genome designations.

However, too much reliance on results from only a few DNA sequences may suffer from the same risks as do other 'one-character phylogenies' (Doyle et al., 1992). Apparently, different genes in the same organism may have different histories due to, for example, bidirectional interlocus concerted evolution (Wendel et al., 1995b), introgression (Wendel and Albert, 1992; Wendel et al., 1995a; Rieseberg and Solits, 1991) or lineage sorting (Pamilo and Nei, 1991). Further, functional dependencies among nucleotide sites may inflate misleading signals (Baldwin et al., 1995). There is no doubt that more and more DNA sequences will be available in the future as the result of current efforts in gene isolation and massive gene sequencing projects.

In the context of elucidating the gene structure of fatty acid desaturases from cotton, I isolated a large intron that is situated in the 5' UTR of the microsomal ω -6 desaturase gene *pghD12-1* of *G. hirsutum*. Because there are few, if any, functional constraints in the evolution of introns, intron sequences are usually evolve more rapidly relative to their flanking exons. Consequently, the sequence data of large introns are often highly phylogenetically informative among closely related taxa. To study the structural and evolutionary aspect of *pghD12-1* gene, I have cloned and sequenced this intron from a range of *Gossypium* species. The alignment and comparison of the intron sequences from allotetraploids and their progenitors, or more precisely the close relatives of their progenitor species, will allow *pghD12-1* gene evolution to be examined on two levels: i.e., in diploids subsequent to their divergence from a common ancestor and in

allotetraploids after the two genomes reunited in a common nucleus. Also of special interest is the phylogenetic relationship among the Australian *Gossypium* species which are among the least understood taxonomic groups.

6.2 MATERIALS AND METHODS

6.2.1 Plant Materials and DNA Extractions

Total genomic DNA was isolated from soybean (*Glycine max*) cv Williams and 32 accessions of 24 *Gossypium* species (Table 6.1) following the protocol described in section 2.4.3 in Chapter Two. Genomic DNA from *A. thaliana* ecotype Colombia was kindly provided by B. Dong (CSIRO Plant Industry).

Table 6.1 A List of *Gossypium* Accessions from Which the pghD12-1 Intron Was Isolated and Sequenced

Species	Taxonomic Authority	Genome	Accession No.	Locality
<i>G. hirsutum</i>	L	AD	Deltapine-16	commercial cultivar
<i>G. barbadense</i>	L	AD	PimaS7	commercial cultivar
<i>G. herbaceum</i>	L	A	LL-1776	Wagad
<i>G. arboreum</i>	L	A	GOS-5264	India Type#9
<i>G. raimondii</i>	Ulbrich	D	CPI136508	Peru
<i>G. trilobum</i>	(DC) Skovsted	D	GOS-5330	Mexico
<i>G. gossypoides</i>	Ulbrich	D	CPI138644	Mexico
<i>G. australe-1</i>	FMueller	G	purchased commercially	Nindethana Seed Service, WA
<i>G. australe-2</i>	FMueller	G	PI-478754	Carawine Gorge (WA)
<i>G. australe -3</i>	FMueller	G	PI-499756	Ormiston Gorge (NT)
<i>G. nelsonii -1</i>	Fryxell	G	GOS-5024	Richmond (QLD)

Table 6.1 con't

<i>G. nelsonii</i> -2	Fryxell	G	PI-499783	Ormiston Gorge (NT)
<i>G. bickii</i> -1	Prokhanov	G	GOS-5338	Urandangie Road S of Barkly Hwy (QLD)
<i>G. bickii</i> -2	Prokhanov	G	PI-464863	Supplejack Station (NT)
<i>G. bickii</i> -3	Prokhanov	G	GOS-5048	Alice Spring (NT)
<i>G. sturtianum</i> -1	JHWillis	C	collected by Peter Abble	10 km to Creek Bed at Hermarinshurg, NT
<i>G. sturtianum</i> -2	JHWillis	C	499810	Glen Helen Station (NT)
<i>G. sturtianum</i> -3	JHWillis	C	5076	Narribri (NSW)
<i>G. robinsonii</i> -1	FMueller	C	purchased commercially	Nindethana Seed Service, WA
<i>G. robinsonii</i> -2	FMueller	C	GOS-5170	Millstream National Park (WA)
<i>G. pulchellum</i>	(CAGardner) Fyxell	K	GOS-5204	Vanstittart Bay (WA)
<i>G. populifolium</i>	(Bentham) GMueller ex Todaro	K	GOS-5210	Augustus Island (WA)
<i>G. costulatum</i>	Todaro	K	GOS-5378	Yampi Peninsula (WA)
<i>G. marchantii</i>	Fryxell, Craven & Stewart	K	GOS-5193	Bougainville Peninsula (WA)
<i>G. londonderriense</i>	Fryxell, Craven & Stewart	K	GOS-5191	Drysdale River Mouth NE of Kalumburu (WA)
<i>G. cunninghamii</i>	Todaro	K	GOS-5309	Cobourg Peninsula (WA)
<i>G. species novum</i> (<i>G. anapoides</i>)	Stewart, Craven and Wendel)	K	GOS-5223	Cape Talbot (WA)
<i>G. nobile</i>	Fryxell, Craven & Stewart	K	GOS-5196	Carson River Station (WA)
<i>G. rotundifolium</i>	Fryxell, Craven & Stewart	K	GOS-5030	16 km North of Broome towards Beagle Bay (WA)
<i>G. pilosum</i>	Fryxell	K	GOS-5203	Northern end of Mitchell Plateau (WA)
<i>G. enthyle</i>	Fryxell, Craven & Stewart	K	GOS-5176	Mitchell River Basin (WA)
<i>G. exiguum</i>	Fryxell, Craven & Stewart	K	GOS-5347	King Edward River (WA)

6.2.2 Adapter-Primer Mediated PCR

In order to isolate the 5' upstream region of the pghD12-1 gene, an adapter-primer mediated PCR procedure was initiated (Espelund and Jakobsen, 1992). Total genomic DNA from *G. hirsutum* cv Deltapine-16 was extracted as described in section 2.4.3 in Chapter Two and a further purification was made by CsCl-EtBr gradients as described in section 2.4.4 in Chapter Two. Approximately 10 µg DNA was digested overnight with EcoRI, and then 50 ng of the EcoRI adapter was ligated to the EcoRI fragments. The nucleotide sequences of the EcoRI adapter are as follows: top strand: 5'-CTCGTAGACTGCGTACC -3'; bottom strand: 5'- AATTGGTACGCAGTC -3'. A hot start PCR was conducted in a 50 µl reaction volume containing 100 ng EcoRI adapter-ligated EcoRI fragments and 10 pmol of each primer. Primer 1 (EcoRI-adapter primer) was the same as the top strand EcoRI adapter, while primer 2 (*d12A4*, see section 5.2.2 in Chapter Five) was designed to anneal to the pghD12-1 approximately 160 bp downstream of the translation initiation site. PCR mixture was preheated to 95°C for 4 min prior to adding Taq polymerase. Following 30 cycles of 94°C for 1 min, 57°C for 1 min and 72°C for 1 min, PCR was extended at 72°C for 10 min. The PCR product was purified using a Wizard PCR Prep column (Promega) and cloned into T[®]easy vector (Promega) according to the manufacturer's instructions. The plasmid DNA of individual clone was isolated using the alkaline/PEG₈₀₀₀ method as described in section 2.4.2 in Chapter Two and DNA sequences were determined using an ABI 373 DNA sequencer.

6.2.3 PCR Amplification and Cloning of 5' UTR Introns from *Gossypium* Species

To amplify the entire pghD12-1 intron revealed by the adapter-primer mediated PCR described above, a set of primers based on the nucleotide sequences flanking the predicted splice site in pghD12-1 were used in PCR reactions with genomic DNAs from *Gossypium* species listed in Table 6.1. The nucleotide sequence of the upstream primer *d12PS1* is: 5'- CCTGGCGTTAAACTGCTTTC -3'. The downstream primer is

d12A4 which was used in the previous adapter-primer mediated PCR reaction. PCRs using other upstream primers including *d12DS1* or *d12it1*, in combination with *d12A4* were also carried out in *G. barbadense* and *G. hirsutum* in order to isolate the D-subgenome intron from these allotetraploids. The *d12DS1* primer was designed on the basis of D-genome-specific DNA sequences of pghD12-1 introns. Its nucleotide sequence is as follows: 5'- AGAATTTAAAATTTAAATCTAACAG -3'. The *d12it-1* primer was designed according to the intron sequence of *G. hirsutum* at 130-150 bp downstream the exon/intron 5' boundary. The PCR conditions and subsequent purification and cloning of the PCR products were the same as described above. Plasmid DNA containing the expected intron and its flanking exon regions was isolated from an individual single colony and prepared using PEG₈₀₀₀ method as described in section 2.4.2 in Chapter Two. DNA sequences were determined using an ABI373 DNA sequencer with seven primers, including forward, reverse, *d12PS1*, *d12A4*, and three other internal primers within the intron to ensure the whole DNA fragments are fully sequenced from both directions.

6.2.4 Sequence Alignment and Analysis

Prior to the multiple sequence analysis, the nucleotides corresponding to the PCR primer regions and the 5' flanking exon region were removed from the sequences. The 32 sequences from 24 cotton species were aligned using the PILEUP module of the Wisconsin GCG computer package, version 8.0 (Devereux et al., 1984). After visual inspection of the resulting alignment, several adjustments in sequence alignment were made. The complete data set was analysed using the GCG program DISTANCES. The observed distances between all pairs of sequences were translated to evolutionarily 'corrected' Kimura two-parameter distances (Kimura, 1980) and subjected to neighbour-joining analysis (Saitou and Nei, 1987).

6.2.5 Amplification of the Corresponding Intron from *A. thaliana* and *G. max*

Based on the published microsomal ω -6 desaturase cDNA sequences, FAD2 from *A. thaliana* and FAD2-1 from *G. max*, one pair of primers was designed for each gene corresponding to the flanking exon regions of the predicted splice sites. The nucleotide sequences of these primers are: *at-1*: 5'- AGAGATTCTGCGGAGGAGC -3' (67 bp upstream the translation initiation site of atFAD2); *at-2*: 5'- GCACGGCACACGC-TTTGTGG -3' (55 bp downstream the translation initiation site of atFAD2); *soy-1* : 5'- CCATATACTAATATTTGCTTGTA -3' (60 bp upstream the translation initiation site of gmFAD2-1); *soy-2*: 5'- AACTTCCACTTTGGCCACAC -3' (37 bp downstream the translation initiation site of gmFAD2-1); The PCR conditions were identical to those described above for the amplification of *Gossypium* introns except that the annealing temperature was adjusted to 56°C for *A. thaliana* and 60°C for *G. max*. The purification of the PCR products and following cloning and sequencing were conducted as described in above section.

6.2.6 High Stringency Southern Blot Analysis

Genomic DNAs of *G. barbadense*, *G. hirsutum*, (cvs Deltapine-16 and Siokra), *G. herbaceum*, *G. raimondii* and *G. robinsonii* were digested by EcoRI and HindIII, respectively and a Southern blot were prepared as described in section 2.4.3 and further purification by CsCl gradient in section 2.4.4 in Chapter Two. A DNA fragment corresponding to the entire 3' UTR region of pghD12-1, which was also used in previously described Southern blot analysis (Chapter Five) was used as a DNA probe. Hybridisation condition was essentially the same as previously described, but the membrane filter was washed at higher stringency conditions. After a brief wash in 2x SSC/0.1% SDS at 65°C, the filter was washed three times in 0.2x SSC/0.1% SDS at 65°C for 20 min each prior to autoradiography.

6.2.7 Construction of a 35S-intron-GUS Cassette

To evaluate the potential enhancing effect of the pghD12-1 intron on gene expression in future experiments, the intron was inserted into a SmaI site between the CaMv35S promoter and GUS (β -glucuronidase) reporter gene. Since this intron is located near the translation initiation codon of pghD12-1, it was necessary to mutate the ATG codon in the 3' flanking exon region so that the expression of the downstream GUS gene would not be affected. A new primer, *it-R*, was designed to mutate the ATG codon to ATA. The DNA sequence of *it-R* primer is as follows: 5'- ATTCCCGGGCACCGGC-ACCTATTTTG -3' (The mutated site is underlined). The pghD12-1 intron was amplified using the *it-R* primer, in combination with *d12PS1* primer. The PCR product was purified using a Wizard PCR prep column (Promega) and cloned in T[®]easy vector (Promega) following the manufacturer's instructions. The plasmid DNA containing the intron insert was isolated from a single colony and digested with NotI (a NotI site appears at each side of the cloning site in T[®]easy vector). The DNA was blunt-ended by adding 1 unit Klenow enzyme and incubated at 37°C for 5 min. Following heat inactivation at 75°C for 10 min, the DNA was purified by phenol/chloroform extraction and ethanol precipitation. In the meanwhile, a binary vector, pBI121 (also shown in Fig. 2.1B in Chapter Two), was digested with SmaI and dephosphorylated by incubation with calf intestinal alkaline phosphatase (Pharmacia) for 30 min at 37°C. Following heat inactivation of the enzyme at 65°C for 10 min, the vector DNA was extracted with phenol/chloroform, chloroform and ethanol precipitated. The DNA pellet was resuspended in 10 μ l dH₂O. A blunt-ended ligation of the intron and SmaI restricted pBI121 was set up at 17°C overnight. The whole ligation was terminated by ethanol precipitation and then electroporated into *E. coli* DH5 α cells prior to growing on the LB agar plates supplemented with 20 μ g/l Ampicillin. Plasmid DNA was prepared from overnight growth of a single colony and sense orientation of the intron was confirmed by DNA sequencing. The plasmid DNA was re-electroporated into *Agrobacterium* and grown on Kanamycin-supplemented LB plates.

6.3 RESULTS

6.3.1 The Localisation of pghD12-1 Intron from *G. hirsutum*

Using an EcoRI adapter-primer mediated PCR, a DNA fragment of approximately 300 bp was amplified from *G. hirsutum* genomic DNA. Fig. 6.1 illustrates the sequence alignment between the genomic PCR fragment and the cDNA clone pghD12-1. As shown in Fig. 6.1, the two sequences were identical until the abrupt dropping of sequence homology at 10 bp upstream the translation initiation site. In the DNA sequence of the PCR fragment-A, a putative 3' intron/exon boundary consensus binucleotide AG was found at the location where the abrupt dropping of sequence homology with pghD12-1 occurred.

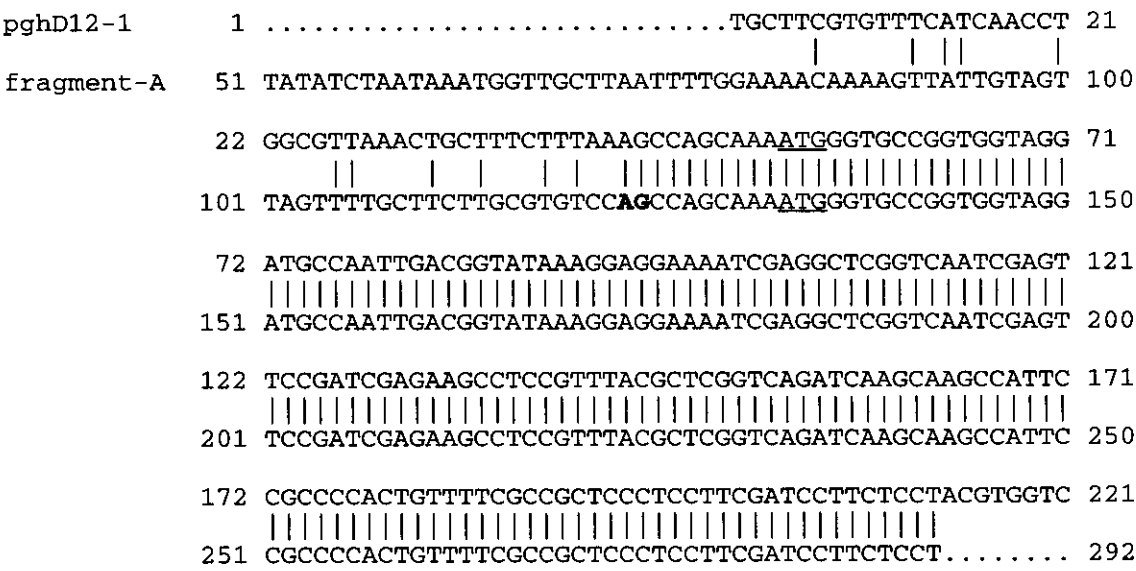


Fig. 6.1 Comparison of nucleotide sequences between the cDNA pghD12-1 and the PCR fragment-A amplified from *G. hirsutum* genomic DNA with primer *d12A4* and the EcoRI primer. The putative 3' intron/exon boundary consensus binucleotide AG is in bold case. The translation initiation site ATG is underlined.

Based on the sequence comparison between cDNA pghD12-1 and the PCR fragment-A derived from *G. hirsutum* genomic DNA, it was predicted that there was an intron within the 5' UTR of pghD12-1. This was confirmed by a further PCR amplification of the entire intron and portions of the flanking exon regions using primers *d12PS1* and *d12A4*. This intron was 1133 bp in size and located at 8 bp upstream of the translation initiation site. This intron starts with binucleotides, GT, and ends with AG, consistent with the plant consensus 5' and 3' exon/intron boundaries (Simpson and Filipowicz, 1996). Further experiments with genomic PCRs by the combinations of various primers covering the entire pghD12-1 cDNA region were conducted to examine the existence of a second or more introns in pghD12-1. However, the sequences of the resulting PCR products matched perfectly with pghD12-1 cDNA (data not shown). Therefore, it is presumed that this large 5' UTR intron is probably the only intron which interrupts the cDNA pghD12-1. This is consistent with the microsomal ω -6 desaturase gene, FAD2, in *A. thaliana* (Okuley et al., 1994). The genomic structure of the seed-specific microsomal ω -6 fatty acid desaturase gene corresponding to the cDNA pghD12-1 from *G. hirsutum* is outlined in Fig. 6.2.

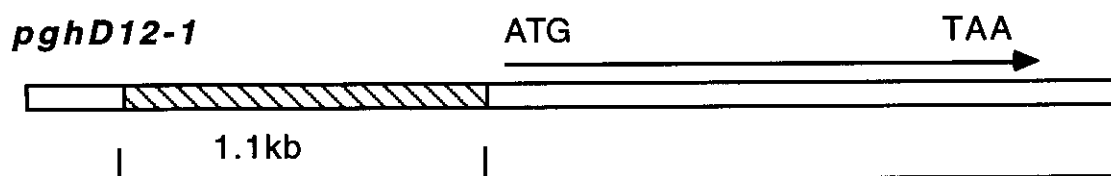


Fig. 6.2 Genomic structure of the microsomal ω -6 desaturase gene corresponding to cDNA pghD12-1 in *G. hirsutum*. The translation initiation site ATG and stop codon TAA are indicated. The exons are represented by open boxes, whereas an 5' UTR intron of 1133 bp was represented by a shaded box.

6.3.2 The Amplification of the pghD12-1 Intron from Other *Gossypium* Species

Using the primer pair, *d12PS1* and *d12A4*, the pghD12-1 intron was obtained from all 32 accessions of the 24 *Gossypium* species listed in Table 6.1. The sequence alignment of the complete intron and a portion of its 3' flanking exon 3' is illustrated in Fig. 6.3. The sequences of primer regions at both sides are removed. Since the 5' flanking exon is very short, it is not included in the analysis. Alignment of all 32 sequences (Fig. 6.3) shows many places of obvious insertions and deletions (indels) which are either species-specific or genome-specific.

Fig. 6.3 Sequence alignment of the 5' UTR intron and a portion of its 3' flanking exon region of pghD12-1 gene in 32 accessions of 24 *Gossypium* species.

*A consensus sequence is drawn by GCG program PRETTY and all nucleotide congruent with consensus is represented by a "-".

*Gaps are represented by dots.

*The 5' and 3' intron/exon boundary is indicated by the binucleotides GT and AG in bold type, respectively.

*The translation initiation site is underlined.

	1					60
trilobum	-----	-----	-----		---g---c---	-----
gossypioides	-----	-----	-----		-----c---	-----
raimondii	-----	-----	-----		---c-----	-----
herbaceum	-----	-----	-----		-----	-----
arboreum	-----	-----t	-----		-----	-----
hirsutum	-----	-----	-----		-----	-----
barbadense	-----	-----	-----		-----	-----
australe-1	-----	-----	-----		-----	-----
australe-3	-----	-----	-----		-----	-----
nelsonii-1	-----	-----	-----c---		-----	-----
nelsonii-2	-----	-----	-----		-----	-----
australe-2	-----	-----	-----		-----	-----
bickii-2	-----	-----c---	-----		-----	-----
bickii-3	-----	-----c---	-----		-----	-----
bickii-1	-----	-----c---	-----		-----	-----
robinsonii-1	-----	-----	-----		-----c---	-----
robinsonii-2	-----	-----	-----		-----	-----
sturtianum-2	-----	-----	-----		-----t---	-----
sturtianum-3	-----	-----	-----		-----	-----
sturtianum-1	-----	-----	-----		-----	-----
londonderriense	-----	-----	-----		-----	-----
anapoides	-----	-----	-----		-----c---	-----
cunninghamii	-----	-----	-----		-----	-----
nobile	-----	-----	-----		-----	-----
marchantii	-----	-----	-----		-----	-----
populifolium	-----	-----	-----		-----t---	-----
costulatum	-----	-----	-----		-----t---	-----
enthyle	-----	-----	-----		-----	-----
rotundifolium	-----	-----	-----		-----	-----
pilosam	-----	-----	-----		-----	-----
exiguum	-----	-----	-----		-----t---	-----
pulchellum	-----	-----	-----		-----t---	-----
CONSENSUS	GTACATTTCT	CTTTAATTC	CTTCATTTTT	TTTTCTTCC	ATTCATGTT	TTCTTGTTA

	61					120
trilobum	-----	-----	-----	-----	-----c---	-----
gossypioides	-----	-----	-----	-----g---	-----	-----
raimondii	-----	-----	-----	-----	-----	-----
herbaceum	-----	-----	-----	-----	-----	-----
arboreum	-----	-----	-----	-----	-----	-----
hirsutum	-----	-----	-----	-----	-----	-----g---
barbadense	-----	-----	-----	-----	-----	-----
australe-1	-----	-----	-----	-----	-----	-----
australe-3	-----	-----	-----	-----	-----	-----
nelsonii-1	-----	-----	-----	-----	-----	-----
nelsonii-2	-----	-----	-----	-----	-----	-----
australe-2	-----c	-----	-----	-----	-----	-----
bickii-2	-----	-----	-----	-----	-----	-----
bickii-3	-----	-----	-----	-----	-----	-----
bickii-1	-----	-----	-----	-----	-----	-----
robinsonii-1	-----	-----	-----	-----	-----c---	-----
robinsonii-2	-----	-----	-----	-----	-----c---	-----
sturtianum-2	-----	-----	-----	-----	-----	-----
sturtianum-3	-----	-----	-----	-----	-----	-----
sturtianum-1	-----	-----	-----t	-----	-----	-----
londonderriense	-----	-----a---	-----	-----	-----	-----
anapoides	-----	-----a---	-----	-----	-----g---	-----c---
cunninghamii	-----	-----a---	-----	-----	-----	-----
nobile	-----	-----a---	-----	-----	-----	-----
marchantii	-----	-----a---	-----	-----	-----g---	-----
populifolium	-----	-----a---	-----	-----	-----	-----
costulatum	-----	-----a---	-----	-----	-----	-----
enthyle	-----	-----a---	-----	-----	-----	-----
rotundifolium	-----	-----a---	-----	-----	-----	-----
pilosam	-----	-----a---	-----	-----	-----	-----
exiguum	-----	-----a---	-----	-----	-----	-----
pulchellum	-----	-----a---	-----	-----	-----	-----
CONSENSUS	ATTTTGCATT	GAAGTGATAA	ATTTGAGTGA	ATGATGTTTG	GTATATCTTC	TTAGTAACTG

	121					180
trilobum	-----	-----		-----	-----	
gossypioides	-----	-----		-----	-----	
raimondii	-----	-----		-----	-----	
herbaceum	-----	-----	-----	-----	-----	
arboreum	-----	-----	-----	-----	-----	
hirsutum	-----	-----	---c-c-	-----	-----	
barbadense	-----	-----	-----	-----	-----	
australe-1	-----	-----		-----	-----	
australe-3	-----	-----		-----	-----	
nelsonii-1	-----	-----		-----	-----	
nelsonii-2	-----	-----		-----	-----	
australe-2	-----	-----		-----	-----	
bickii-2	-----	-----		-----	-----	
bickii-3	-----	-----		-----	-----	
bickii-1	-----	-----		-----t-	-----	
robinsonii-1	-----	-----		-----t-	-----	
robinsonii-2	-----	-----		-----	-----	
sturtianum-2	-----	-----		-----	-----	
sturtianum-3	-----	-----		-----	-----	
sturtianum-1	-----c-	-----		-----	-----	
londonderriense	g-----	-----	..c..	-----	-----	
anapoides	-----	-----		-----	-----	
cunninghamii	-----	-----		-----	-----	
nobile	-----c-	-----		-----	-----	
marchantii	-----	-----		-----	-----	
populifolium	-----	-----		-----	-----aaag-	-----
costulatum	-----	-----		-----	-----	
enthyle	-----	-----		-----	-----	
rotundifolium	-----	-----		-----	-----	
pilosam	-----	-----t-		-----	-----	
exiguum	-----	-----		-----	-----	
pulchellum	-----	-----		-----	-----	
CONSENSUS	ACCTTTTGAA	AATACTAGCA	TTTTTTTTTT	TTAATATCAA	GTGA---AA	GAAGAAGAAG

	181					240
trilobum	...-----	---t----	-----t--	-----ag	-----	...
gossypioides	...-----	---gt----	-----	-----	-----	...
raimondii			-----	-----	-----	...
herbaceum	-----	-----	-----	-----	-----	...
arboreum	-----	-----	-----g	-----	-----	...
hirsutum	g-----	-----		-----	-----	...
barbadense	-----	-----		-----	-----	...
australe-1	-----	-----	a-g	-----	-----	...
australe-3	-----	-----	-----	-----	-----	...
nelsonii-1	---.....			-----	-----	...
nelsonii-2	---.....			-----	-----	...
australe-2	---.....			-----	-----	...
bickii-2			-----	-----	-----	...
bickii-3			-----	-----	-----	...
bickii-1			-----	-----	-----	...
robinsonii-1	-----	-----	-----	-----	-----	...
robinsonii-2	-----	-----	-----	-----	-----	...
sturtianum-2		-----	-----	-----t-	-----	...
sturtianum-3		-----	-----	-----	-----	...
sturtianum-1		-----	-----	-----	-----	...
londonderriense	-----	-----	-----	-----	-----	...
anapoides	...-----	-----	-c-	-----	-----	...
cunninghamii	-----	-----	-----	-----	-----	...
nobile	...-----	-----c-	-----	-----	-----	...
marchantii	-----	-----	-----	-----	-----	...
populifolium	-----	-----	-----	-----t-	-----	...
costulatum	...-----	-----	-c-	-----	-----	...
enthyle	-----	-----	-c-	-----	-----	...
rotundifolium	-----	-----g	-----	-----	-----	...
pilosam	-----	-----	-----	-----	-----	...
exiguum	-----	-----g	-----	-----	-----	...
pulchellum	-----tc-	-----	-----	-----	-----	...
CONSENSUS	AAGAAGAAGA	AGAAGAAGAA	GAAGAAGAAT	TCGCCATGC	AAAAAGCTTT	TTAAGGCTTT

	241					300
trilobum	-c	-c-t				
gossypioides	-c	-c-t				
raimondii	-c	-c-t				
herbaceum	-c	-t	-g			
arboreum		-t	-g	-c		
hirsutum		-t				
barbadense		-t				-g
australe-1	-g	C				
australe-3	-g	C				
nelsonii-1	-g	C		-c		
nelsonii-2	-g	C		-c		
australe-2	-g	C				
bickii-2		C				
bickii-3		C				
bickii-1		C				
robinsonii-1	c	-ta				
robinsonii-2	c					
sturtianum-2	c	-g				
sturtianum-3	c	-g				
sturtianum-1	c	-g				
londonderriense		-c				
anapoides	a					
cunninghamii				tt		
nobile						
marchantii						
populifolium		-c				
costulatum						-t
enthyle						
rotundifolium				-c		
pilosam						
exiguum						
pulchellum		-c	-t			
CONSENSUS	TTTTTTTTTT	TTTTTCATAG	ATCAAAATTT	ATTTGTTTAC	TTATACTGTT	CTTTTAAGCC

	301					360
trilobum						-t
gossypioides						
raimondii					-a	
herbaceum			tt	a	-g	
arboreum			tt	a		
hirsutum			tt	a		
barbadense			tt	a		
australe-1						
australe-3						
nelsonii-1						
nelsonii-2						
australe-2						
bickii-2						
bickii-3						
bickii-1						
robinsonii-1						
robinsonii-2						
sturtianum-2		a				
sturtianum-3		a				
sturtianum-1		a				
londonderriense						-g
anapoides						
cunninghamii				c		
nobile						
marchantii		c				
populifolium						
costulatum						
enthyle						
rotundifolium						
pilosam						
exiguum				g		
pulchellum		c	a	t	g	
CONSENSUS	TGAAGAAAGA	AGCCATGGTT	TCAATTTTGT	AGAGAGTTT	AAATCCCAGA	TACCAGAGAG

	361					420
trilobum	-----	-----	-----	-----	-----	-----
gossypioides	--c-----	-----	-----	-----	-----	-----
raimondii	-----	-----	-----	-----	-----	-----
herbaceum	-----a--	-----	-----g--	-----	-----	-----t--
arboreum	-----	-----	-----g--	-----	-----	-----t--
hirsutum	-----	-----	-----g--	-----	-----	-----t--
barbadense	-----	-----	-----g--	-----	-----	-----
australe-1	-----	-----	-----	-----	-----	-----
australe-3	-----	-----	-----	-----	-----	-----
nelsonii-1	-----	-----	-----a--	-----	-----	-----
nelsonii-2	-----	-----	-----a--	-----	-----	-----
australe-2	-----	-----	-----	-----	-----	-----
bickii-2	-----	-----	-----	-----	-----	-----
bickii-3	-----	-----	-----	-----	-----	-----
bickii-1	-----	-----	-----	-----c-----	-----	-----
robinsonii-1	-----	-----c-----c-----	-----	-----	-----	-----
robinsonii-2	-----	-----c-----c-----	-----	-----	-----	-----
sturtianum-2	-----	-----	-----	-----	-----	-----
sturtianum-3	-----	-----	-----	-----	-----	-----
sturtianum-1	-----	-----	-----	-----	-----	-----
londonderriense	-----	-----	-----	-----	-----t-----	-----
anapoides	-----	-----	-----	-----	-----t-----	-----
cunninghamii	-----	-----	-----t-----	-----	-----t-----	-----
nobile	--g-----	-----	-----	-----	-----t-----	-----
marchantii	-----	-----	-----	-----	-----t-----	-----
populifolium	-----	-----	-----	-----	-----t-----	-----
costulatum	-----	-----	-----	-----	-----t-----	-----
enthyle	-----	-----	-----	-----	-----	-----
rotundifolium	-----	-----	-----	-----	-----	-----
pilosam	-----	-----	-----c-----	-----	-----t-----	-----
exiguum	-----	-----	-----	-----	-----	-----
pulchellum	-----	-----	-----	-----	-----t-----	-----
CONSENSUS	CTTCATCGTT	TATTCATATA	TTTTTAAACA	TTTTTTAAAG	CAAGAAGCTG	TGATTTTGGT

	421					480
trilobum	-----	-----	-----t-----	-----	-----	-----t-----
gossypioides	-----t-----	-----	-----t-----	-----	-----	-----t-----
raimondii	-----	-----	-----t-----	-----	-----	-----g--t-----
herbaceum	-----	-----	-----	-----	-----	-----
arboreum	-----	-----	-----	-----	-----	-----
hirsutum	-----	-----	-----	-----	-----	-----
barbadense	-----	-----	-----	-----	-----	-----
australe-1	-----	-----c-----	-----	-----c-----	-----	-----
australe-3	-----	-----	-----	-----	-----	-----
nelsonii-1	-----	-----	-----	-----	-----	-----
nelsonii-2	-----	-----	-----	-----	-----	-----
australe-2	-----	-----	-----	-----	-----	-----
bickii-2	-----	-----	-----	-----	-----	-----
bickii-3	-----	-----	-----	-----	-----	-----
bickii-1	-----	-----	-----	-----	-----	-----
robinsonii-1	-----	-----	-----	-----t-----	-----	-----
robinsonii-2	-----c-----	-----	-----	-----t-----	-----	-----
sturtianum-2	-----	-----	-----a-----	-----	-----	-----
sturtianum-3	-----	-----	-----a-----	-----	-----	-----
sturtianum-1	-----	-----	-----a-----	-----t-----	-----	-----
londonderriense	-----g-----	-----	-----c-----	-----a-----	-----	-----
anapoides	-----c-g-----	-----	-----c-----	-----a-----	-----	-----
cunninghamii	-----g-----	-----	-----c-----	-----a-----	-----	-----g-----
nobile	-----g-----	-----	-----a-----c-----	-----a-----	-----	-----
marchantii	-----g-----	-----	-----c-----	-----a-----	-----	-----
populifolium	-----g-----	-----	-----c-----	-----a-----	-----	-----
costulatum	-----g-----	-----	-----	-----a-----	-----	-----
enthyle	-----	-----	-----t-----	-----	-----	-----
rotundifolium	-----a-----	-----	-----	-----	-----	-----
pilosam	-----g-----	-----t-----	-----	-----	-----	-----
exiguum	-----	-----	-----	-----	-----	-----
pulchellum	-----	-----	-----c-----	-----a-----	-----	-----
CONSENSUS	TTTGCTAAAA	TATGCAATAA	ATCTTCATAT	TTTTCGTAAA	TTTAAAAATTT	AATCCTTTTAA

481 540

trilobum	ct---aga-t	--a-g-t---	act-----	-----	--t--....	
gossypoides	ct---aga-t	--a--t---	ac-----	-----	--t--cg-t-	-a-----
raimondii	ct---a---t	--a---t---	act-----	-----	--t--....	
herbaceum	gt-----	-----		-a--t	ctacgt-aa-	t---a-.-
arboreum	gt-----	-----		-a--t	ctacgt-aa-	t---a-.-
hirsutum	gt-----	-----		-a--t	ctacgt-aa-	t---a-.-
barbadense	gt-----	-----		-a--t	ctacgt-aa-	t---a-.-
australe-1	-----	-----t---	-----	-----	-----	-----
australe-3	-----	-----ct---	-----	-----	-----	-----
nelsonii-1	-----g	-----	-----ag	-----	-----	-----
nelsonii-2	-----g	-----	-----ag	-----	-----	-----
australe-2	-----	-----	-----ag	-----	-----	-----
bickii-2	-----c	-----	-----ag	at---.--t	acacg.---	-----
bickii-3	-----c	-----	-----ag	at---.--t	acacg.---	-----
bickii-1	-----c	-----	-----ag	at---.--t	acacg.---	-----
robinsonii-1	-----	-----	-----ag	at--a--t	aca.ga---	-----
robinsonii-2	-----	-----	-----ag	at--a--t	aca.ga---	-----
sturtianum-2	-----	-----	-----ag	at--a--t	acacg.---	-----
sturtianum-3	-----	-----	-----ag	at--a--t	acacg.---	-----
sturtianum-1	-----	-----	-----ag	at--a--t	acacgt---	-----
londonderriense	-----	-----,c-	-tac-----	-----	--c---a-	-----c--
anapoides	-----	-----	-tac-----	-----	-gc---a-	-----
cunninghamii	-----	-----	-tac-----	-----	-c--a--a-	-----
nobile	-----	-----,c-	-tac-----	-----	-c---a-	-----
marchantii	-----	-----,c-	-tac-----	-----	-c---a-	-----
populifolium	-,-----	-----	-----	-----	-c---...	
costulatum	-----c--	-----	-----	t-----	-c--a...	
enthyle	-----a--	-----	-----	-----	-c-----	-----
rotundifolium	-----a--	-----	-----	a-----	-c-----	-----
pilosam	-----a--	-----	-----	-----	-c-----	-----
exiguum	-----a--	-----	-----	-----	-c--a---	-----
pulchellum	-----	-----	-----	-----	-c-----	-----
CONSENSUS	TCCTTTTAGAA	TTTAATCTTT	TACTTTTTTA	GATTTTAAAA	TA-ACGTTGA	ATTTTTAACA

541 600

trilobum		.g-----	--,-----	-----a---	-----a.--	-----
gossypoides		-----	-----	-----a---	-----	-----g-
raimondii		.g--c---	--,--a---	-----a---	-----	-----a-
herbaceum	--a-----	-g-----	-----	-----t---	-----...t	a-----
arboreum	--a-----	-----	-----	-----t---	-----...t	a-----
hirsutum	--a-----	-----	-----	-----t---	-----...t	a-----
barbadense	--a-----	-----	g-----	-----t---	-----...t	a--c-----
australe-1			..-----	-----	-----,at	-----
australe-3			..-----	-----	-----,t	-----
nelsonii-1			-----	-----	-----,a.	-----g-
nelsonii-2			-----	-----	-----,a.	-----
australe-2			-----	-----	-----,at	-----
bickii-2	-----	..-----	-----	-----,t	-a-t,-g-	-----
bickii-3	-----	..-----	-----	-----,t	-a-t,-g-	-----
bickii-1	-----	..-----	-----	-----,t	-a-t,-g-	-----
robinsonii-1	-----t-	-----	-----	-----,t	-a-t,-g-	-----
robinsonii-2	-----t-	-----	-----	-----,t	-a-t,-g-	-----
sturtianum-2	-----	-----	-----	-----,t	-a-t,-g-	-----
sturtianum-3	-----	-----	-----	-----,t	ga-t,-gg-	-----
sturtianum-1	-----	-----	-----	-----,t	-a--,-g-	-----
londonderriense	--c-----	-----	-----	-----	-----,.	-----c-----
anapoides	--c-----	-----	-----	-----a-	-----,.	-----
cunninghamii	--c-----g-	-----	-----	-----	-----,a	-----
nobile	--c-----	-----	-----	-----	-----g-t	-----
marchantii	--c-----	-----	-----	-----	-----	-----
populifolium	-a.....		-----	t-----a-	-----,a.	-----
costulatum	-a.....		-----	-----a-	-----,a.	-----
enthyle	-----		-----	-----	-----,a.	-----
rotundifolium	-----		-----	-----	-----,a.	-----
pilosam	-----		-----	-----	-----,a.	-----
exiguum	-----		-----	-----	-----,at	-----
pulchellum	-----		-----	-----	-----,a.	-----
CONSENSUS	TTGTTAAAT	TATTTTATTA	AATTTAGATT	ATTATGAGTA	ATTGAGATTA	TTTTTTTTAT

	601					660
trilobum	-----	---t-----	-----	-----	-----	-----
gossypioides	---t-----	c-t-----	---t-----	-----	-----	-----
raimondii	---t-----	c-----	---t-----	-----	-----	-----
herbaceum	-----	-----	---t-----	-----	-----	-----
arboreum	-----	-----	---t-----	-----	-----	-----
hirsutum	-----	-----	---t-----	-----	-----	-----
barbadense	-----	-----	---t-----	-----	-----	-----
australe-1	-----	-----	-----	-----	-----	-----
australe-3	-----	-----	-----	-----	-----	-----
nelsonii-1	-----	-----	-----	-----	-----	-----
nelsonii-2	-----	-----	g-----	-----	-----	-----
australe-2	-----	-----	-----	-----	-----	-----
bickii-2	-----	-----	-----	-----	-----	-----
bickii-3	-----	-----	-----	-----	g-----	-----
bickii-1	-----	-----	-----	-----	-----	-----
robinsonii-1	-----	-----	g-----	-----	-----	-----
robinsonii-2	-----	-----	-----	-----	-----	-----
sturtianum-2	-----	-----	-----	-----	-----	-----
sturtianum-3	-----	-----	-----	-----	-----	-----
sturtianum-1	-----	-----	-----	-----	-----	-----
londonderriense	-----	-----	-----	-----	-----	a-----
anapoides	-----	-----	aa-----	-----	-----	a-----
cunninghamii	-----	-----	-----	-----	-----	a-----
nobile	-----	-----	-----	-----	-----	-----
marchantii	-----	-----	-----	-----	-----	-----
populifolium	-----	-----	-----	-----	-----	-----
costulatum	-----	-----	-----	-----	-----	g-----
enthyle	-----c-----	-----	-----	-----	-----	a-----
rotundifolium	-----c-----	-----	-----	-----	-----	a-----
pilosam	-----c-----	-----	-----	-----	-----	a-----
exiguum	---c-----c-----	-----	-----	-----	-----	a-----
pulchellum	-----	-----	-----	-----	-----	-----
CONSENSUS	TTCAAAATTT	TACATCATCA	AAGT--AAAA	AAAAAAAAAAC	AAAGTTAACA	CTTGTTACTTG

	661					720
trilobum	---tg-g-a	g-a--g--	-----	-----a--	a---ac---	---c-----
gossypioides	---tg-a-a	g-a--g--	-----	-----g-	a---ac---	---c-----
raimondii	-.....	-t	-----c-	-----t-g-	a---a---	---c-----
herbaceum	-----	-a---a..		-	-a-c-t-	-----
arboreum	-----	-a---a..		-	-a-c-t-	---a-----
hirsutum	-----	-a---a..		-	-a-c-t-	-----
barbadense	-----	-a---a..		-	-a-c-t-	-----
australe-1	-----	-----c..		-	-t-----	-----
australe-3	-----	-----c..		-	-t-----	-----
nelsonii-1	-----	-----c..		-	-t-----	t-----
nelsonii-2	-----	-----c..		-	-t-----	t-----
australe-2	-----	g--g--c..		-	-t-----	t-----
bickii-2	---t-----	-----		-	-----c-	-----
bickii-3	---t-----	-----		-	-----c-	-----
bickii-1	---t-----	-----		-	-----c-	-----
robinsonii-1	---t-----	-----		-	-----c-	-----
robinsonii-2	---t-----	-----		-	-----c-	-----
sturtianum-2	-----	-----		-	-----c-	---a-----
sturtianum-3	-----	-----		-	-----c-	---a-----
sturtianum-1	-----	-----		-	-----c-	---a-----
londonderriense	-----t-	-----		-	-----	-----
anapoides	-----	-----t..		-	-----	-----
cunninghamii	-----	-----		-	-----	-----
nobile	-----	-----		-	-----	-----
marchantii	-----	-----		-	-----	-----
populifolium	-----c-----	-----		-	a-----	-----
costulatum	-----	-a-----		-	-----	-----
enthyle	-----	-a-----		-	-g-----	-----
rotundifolium	-----	-----		-	-----	-----
pilosam	-----	-a-----		-	a-----	-----
exiguum	-----	-a-----		-a-	-----	-----
pulchellum	-----	-----		-----	-----	-----
CONSENSUS	ATTCAATAT	CTGAAAAGAC	TAAATTTTAA	ATTTTCATTA	GAAGGGTTAA	ATTTTAAATT

	721					780
trilobum	-----	---a-c-	t---ca---	-agat-----	-a-----	---g---
gossypoides	-----	--c-a-c-	---ca---	-agat-----	-a-----	-g--g---
raimondii	-----	---a-c-	---ca---	-aga-----	-a-----	---g---
herbaceum	-----	-----	---c---	-----	g-----	-----
arboreum	---g---	-----	---c---	-----	g-----	-----
hirsutum	-----	-----	---c---	-----	g-----	-----
barbadense	-----	-----	---c---	-----	g-----	-----
australe-1	--.....		---c---	-----	-----	-----
australe-3	--.....		---c---	-----	-----	-----
nelsonii-1	--.....		---c---	-----	---c-----	-----
nelsonii-2	--.....		---c---	-----	---c-----	-----
australe-2	--.....		---c---	-----	-----	-----
bickii-2	-----	-----	-----	---c-----	-----	-----
bickii-3	-----	-----	-----	---c-----	-----	-----
bickii-1	-----	-----	-----	---c-----	-----	-----
robinsonii-1	-----	-----	---c---	-----	-----	-----
robinsonii-2	-----	-----	---c---	-----	-----	-----
sturtianum-2	-----	-----	---c---	-----	-----	-----
sturtianum-3	-----	-----	---c---	-----	-----	-----
sturtianum-1	-----	-----	---c---	-----	-----	-----
londonderriense	-----	---a---	t---,---	-----	-----	-----
anapoides	-----	---a---	t---,---	-----	-----	-----
cunninghamii	-----	---a---	t---,---	-----t-	-----	-----
nobile	-----	g-a---	t---,---	-----	-----	-----
marchantii	-----	---a---	t---,---	-----	-----	-----
populifolium	-----	---a---	t-a---,---	-----	-----	-----
costulatum	-----	---a---	t-a---,---	-----	-----	-----
enthyle	-----	-----	t---,---	-----	-----	-----
rotundifolium	-----	---a---	t---,---	-----	-----	-----
pilosam	-----	-----	t---,---	-----	-----	-----
exiguum	-----	---a---	t---,---	-----	-----	-----
pulchellum	-----	---ac---	t-a---,---	-----	-----	-----
CONSENSUS	TTCAAAGAGT	ATAGGGATT	CTCTAATTTT	TTAGATTTA	ATTTTAAATC	TAACAATTAA

	781					840
trilobum	-----	---t-ag	ga-----	--,.....		
gossypoides	-----	---t-a-	ga-----	--,.....		
raimondii	-----	---t-a-	ga-----	--,.....	---g---	---
herbaceum	-----	-----	-----	--,.....	---g---	---
arboreum	-----	-----	-----	--,.....	---g---	---
hirsutum	-----	-----	-----	--,.....	-----	---
barbadense	-----	-----	-----	--,.....	-----	---
australe-1	g-----	-----	-----	--,.....	---t---	---
australe-3	g-----	-----	-----	--,.....	---t---	---
nelsonii-1	g-----	-----	-----	--,.....	---t---	---
nelsonii-2	g-----	-----	-----	--,.....	---t---	---
australe-2	g-----	-----	-----	--,.....	---t---	---
bickii-2	t-----	-----	-----	--,.....	-----	---
bickii-3	t-----	-----	-----	--,.....	-----	---
bickii-1	t-----	g-----	-----	--,.....	-----	---
robinsonii-1	-----	-----	-----	--,.....	-----	---
robinsonii-2	-----	-----	-----	--,.....	-----	---
sturtianum-2	-----	-----	-----	--,.....	-----	---
sturtianum-3	-----	-----	-----	--,.....	-----	---
sturtianum-1	-----	-----	-----	--,.....	-----	---
londonderriense	-----t-	-----	-----	--,.....	-----	---
anapoides	-----t-	-----	---c---	--,.....	-----	---
cunninghamii	-----t-	-----	-----	--,.....	---g---	---
nobile	-----t-	-----	-----	--,.....	-----	---
marchantii	-----t-	-----	-----	--,.....	---t---	---
populifolium	---,--t-	-----	-----	--,.....	-----	---
costulatum	---,--t-	-----	-----	--,.....	-----	---
enthyle	-----t-	-----t	---a---	-----	-----	---
rotundifolium	-----t-	-----	---g-g-	-----	-----	---
pilosam	-----t-	-----	-----	--,.....	-----	---
exiguum	-----t-	-----	-----	--,.....	-----	---
pulchellum	-----t-	-----	-----	--,.....	-----	---
CONSENSUS	CACCTTCTTA	ATTACATTTT	ACTAAATTTA	ACT-AAAAAA	TTACAATATT	AATAGTTAAA

	841					900
trilobum		-a--a--t-	----t----	-g-----	-----	-----
gossypoides	-----	aa--c--gt-	----t----	-g-----	-----	-----
raimondii	-----	-a-----	-a-----	-g-----	-g-----	-----
herbaceum	-----	a-----	-a-----	-----	-----	-a-----
arboreum	-----	a-----	-a-----	-----	-----	-a-----
hirsutum	-----	-----	-a-----	-----	-----	-a-----
barbadense	-----	-----			-----	-a-----
australe-1	-----	-----	-----	-----	-----	-----c
australe-3	-----	-----	-----	-----	-----	-----c
nelsonii-1	-----	-a-----	-----	-----	-----	-----c
nelsonii-2	-----	-a-----	-----	-----	-----	-----c
australe-2	-----	-----	-----	-----	-----	-----c
bickii-2	-----	-a-----	-----	-----	-----	-----
bickii-3	-----	-a-----	-----	-----	-----	-----
bickii-1	-----	-a-----	-----	-----	-----	-----
robinsonii-1	-----	-a-----	-a-----	-----	-----	-----
robinsonii-2	-----	-a-----	-ac-----	-----	-----	-----
sturtianum-2	-----	-----	-----t-	-t-----	-----	-----
sturtianum-3	-----	-----	-----t-	-t-----	-----	-----
sturtianum-1	-----	-----	-----t-	-t-----	-----	-----
londonderriense	-----	-----	-----	-----	-a-----	-----c
anapoides	-----	-----	-----g-	-----	-a-----	-----c
cunninghamii	-----	-----	-----	-----	-a-----	-----g-----c
nobile	-a-----	a-----	-----	-----	-a-----	-----
marchantii	-----	-----	-----	-----	-a-----	-----
populifolium	-----	-----	-----	-----	-a-----	-----
costulatum	-----	-----	-----	-----	-a-----	-----g-----
enthyle	-----	-----	-----	-a-----	-----	-----
rotundifolium	-----	-----	-----	-a-----	-----	-----
pilosam	-----	-----	-----	-a-----	-----	-----
exiguum	-----	-----	-----	-a-----	-----	-----
pulchellum	-----	-a-----	-a-----	-----	-a-----	-----
CONSENSUS	ATTAAATTTT	GGAAAGTACA	AGGACTAAAT	TTTAAATTTT	CAAAAAGTAT	AGGAAGTTAT

	901					960
trilobum	----c----	-----	-----	c-----	-----	-----t-
gossypoides	-----	-----	-----	c-----	-----	-----t-
raimondii	-----	-----	-----c	c-----	-a-----	-----t-
herbaceum	----c----	-----	-----c	-----	-----	-----
arboreum	----c----	-----	-g-----	c-----	-----	-----
hirsutum	----c----	-----	-----	c-----	-----	-----
barbadense	----ct----	-----	-----	c-----	-----	-----
australe-1	-----	-----	-----	-----	-----	-----
australe-3	-----	-----	-----	-----	-----	-----
nelsonii-1	-----	-----	-----g-	-----	-----	-----
nelsonii-2	-----	-----	-----g-	-----	-----	-----
australe-2	-----	-----	-----	-----	-----	-----
bickii-2	-----	-----	-----	-----g-	-----	-----c-----
bickii-3	-----	-----	-----	-----	-----	-----c-----
bickii-1	-----	-----	-----	-----	-----	-----c-----
robinsonii-1	-----	-----	-----	-----	-----	-a-----g-
robinsonii-2	-----	-----	-----	-----	-----	-a-----g-
sturtianum-2	-----	-----	-----	-----g	-----	-----
sturtianum-3	-----	-----	-----	-----g	-----	-----
sturtianum-1	-----	-----	-----	-----g	-----	-----
londonderriense	-----	-----	-----c-----	-a-----	-----	-a-----
anapoides	-----	-----	-----	-----	-----	-a-----
cunninghamii	-----	-----	-----	-----	-----	-a-----
nobile	-----	-----	-----	-----	-t-----	
marchantii	-----	-g-----	-----	-gatga-g	-t-c-----	-----
populifolium	-----	-----	-----	-----	-c-----	-a-----
costulatum	-----	-----	-----	-----	-c-----	-a-----g-----
enthyle	-----	-----	-----	-----c	-a-----	-a-----
rotundifolium	-----	-----	-----	-----c	-----	-----
pilosam	-----	-----	-----	-----c	-----	-a-----
exiguum	-----	-----	-----	-----c	-----	-a-----
pulchellum	-----	-----	-----	-----	-t-----	-t-----
CONSENSUS	AGTATATTTT	AACCTTTATT	TATTTTATAT	ATGATGAGGT	TCCTGCATGC	ACCGAAGAAG

	961					1020
trilobum	-----c--	-a--	---	-c-----	----t-t-	-----
gossypioides	-----	-a--t-	---	-c-----	-----	-----
raimondii	-----	-a--.	---	-c-----	----g--	-----
herbaceum	-----	-a--	---	-----	-----	-----
arboreum	-----	-a--	---	-a-----	-----	-----
hirsutum	-----	-a--.	---	-----	----t-t-	-----
barbadense	-----	-a--	---	-----	a-----	-----
australe-1	-----	-a--.	---	-a-----	-----	-----
australe-3	-----	-a--.	---	-a-----	-----	-----
nelsonii-1	-----	-a--	---	-a-----	-----	-----
nelsonii-2	-----	-a--.	---	-a-----	-----	-----
australe-2	-----	-a--.	---	-a-----	-----	-----
bickii-2	-----	-a--	---	-----	----cg-	-----
bickii-3	-----	-a--	---	-----	----g-	-----
bickii-1	-----	-a--.	---	-----	----g-	-----
robinsonii-1	-----	-----	---	c-----	-----	-----
robinsonii-2	-----	-----	---	-----	-----	-----
sturtianum-2	-----	-----	---	-----	-----	-----
sturtianum-3	-----	-----	---	-----	-----	-----
sturtianum-1	-----	-----	---	-----	-----	-----
londonderriense	-----c-	-----	---	-----	-----	-----
anapoides	----g--c-	-----	---	-----	-----	-----
cunninghamii	-----c-	-----	---	-----	-----	-----
nobile	----t-----	-----	---	-----	-----	-----
marchantii	-----.	-----	---	-----	-----	-----
populifolium	-----	-----	---	-----	-----	-----t
costulatum	-----	-----	---	-----	-----	-----
enthyle	-----	-----	---	-----	-----	-----g-
rotundifolium	-----	-----	---	-a-----	-----	-----
pilosam	-----	-----	---	-----	-----	-----
exiguum	-----	-----	---	-----	----t-t-	-----
pulchellum	-----	-----	---	-----	-----	-----
CONSENSUS	TCACCTTTTG	CCGGTA-TTT	TCGAGTGGCT	TGTTTCTCTC	AAAACTACCT	TGAATCTTGA

	1021					1080
trilobum	-----	-----	---	..-----	-----	-----
gossypioides	-----	-----	---	..-----	-----	-----
raimondii	-----	-----c-	---	..-t-	-----	-----
herbaceum	-----	-----a-	---	..-----	-----	-----c-
arboreum	-----	-----a-	---	..-----	-----	-----c-
hirsutum	-----	-----a-	---	..-----	-----	-----
barbadense	-----	-----a-	---	..-----	-----	-----
australe-1	---.-g--	-----	---	----tt-	-----	-----
australe-3	---.-g--	-----	---	----tt-	-----	-----
nelsonii-1	---.-g--	-----	---	---.t-	c-----	-----
nelsonii-2	---.-g--	-----	---	---.t-	-----	-----
australe-2	---.-g--	-----	---	---.t-	-----t-	-----
bickii-2	---g-----	-----	---	---.-g-	-----	-----
bickii-3	---g-----	-----	---	---.-g-	-----	-----
bickii-1	---g-----	-----	---	---.-g-	-----	-----
robinsonii-1	-----	-----	---	---.-g-	-----	-----
robinsonii-2	-----	-----	---	---.-g-	-----	-----
sturtianum-2	-----	-----	---	---.-g-	-----	-----
sturtianum-3	-----	-----	---	---.-g-	c-----	-----
sturtianum-1	-----	-----	---	---.-g-	-----	-----
londonderriense	-----	-----	---	---.-g-	-----	-----
anapoides	-----	-----	---	---.-g-	-----	-----
cunninghamii	-----	-----	---	---.-g-	-----	-----
nobile	-----	-----	---	---.-g-	-----	-----
marchantii	-----	-----	---	---.-g-	-----	-----
populifolium	-----	-----	---	---.-g-	-----	-----
costulatum	---.-g--	-----	---	---.-g-	-----	-----
enthyle	-----	-----	---	---.-g-	-----c-	-----
rotundifolium	-----	-----	---	---.-g-	-----	-----
pilosam	---c-----	-----	---	---.-g-	-----	-----
exiguum	-----	-----	---	---.-g-	-----	-----
pulchellum	-----	-----	---	---.-g-	-----	-----
CONSENSUS	GACAGAAATTA	AATATATTTT	TGGCCTTTTC	TTCCTAATTT	TTCTCTCTCT	ATTTTCTTTT

	1081					1140
trilobum	--g-----	-----	-----	-----	-----	-----t-----
gossypioides	-----	-----	-----	-----	-----	-----t-----
raimondii	t-----	-----	-----	-----	-----	-----t-----
herbaceum	-----	-----	-----	-----	-----	-----
arboreum	-----a--	-----	-----	-----	-----	-----c-----
hirsutum	-----	-----	-----	-----	-----	-----
barbadense	-----	-----	-----	-----	-----	-----
australe-1	-----	-----	-----	-----	-----g--	-----
australe-3	-----	-----	-----	-----	-----g--	-----
nelsonii-1	.-----	-----	-----	-----g	-----	-----
nelsonii-2	-----c--	-----	-----	-----g	-----	-----
australe-2	-----	-----	-----	-----g	-----	-----
bickii-2	-----a--	-----	-----	-----	-----a--	-----
bickii-3	-----	-----	-----	-----	-----a--	-----
bickii-1	-----	-----	-----	-----	-----aa--	-----
robinsonii-1	-----	-----	-----	-----	-----	-----
robinsonii-2	-----	-----	-----	-----	-----	-----
sturtianum-2	-----	-----	-----	-----	-----	-----
sturtianum-3	-----	-----	-----	-----	-----	-----
sturtianum-1	-----	-----	-----	-----	-----	-----
londonderriense	-----a--	-----	-----	-----	-----a--	-----
anapoides	-----	-----	-----t-	-----	-----a--	-----
cunninghamii	-----	-----	-----	-----	-----a--	-----
nobile	-----.	-----	-----	-----	-----a--	-----c
marchantii	-----a--	-----	-----	-----	-----a--	-----
populifolium	-----	-----	-----	-----	-----a--	-----
costulatum	-----	-----	-----	-----	-----a--	-----
enthyle	-----g--	-----	-----	-----	-----a--	-----
rotundifolium	-----	-----	-----	-----	-----a--	-----
pilosam	-----	-----	-----c--	-----	-----a--	-----
exiguum	-----	-----	-----	-----	-----a--	-----
pulchellum	-----a--	-----	-----	-----	-----a--	-----
CONSENSUS	AAAAATTGCT	TTAGAGAATT	CAGAAAAAAT	ACTTTCCAAC	ACGAAAATTG	CTTCAAATTT

	1141					1200
trilobum	-----	-----	-----	---t-.g--	-c-----	-----g--
gossypioides	-----	-----	-----	---t-.g--	-c-----	-----g--
raimondii	-----	-----	-----	---t-.g--	-c-----	-----g--
herbaceum	--g-----	-----	-----	---.	-----	-----
arboreum	--g-----	-----	-----	---.	-----	-----
hirsutum	--g-.-----	-----	-----	---.	-----	-----
barbadense	--g-----	-----	-----	---.	-----	-----
australe-1	-----	-----	-----	---.	-----	-----
australe-3	-----	-----	-----	---.	-----	-----
nelsonii-1	-----	-----	-----	---.	-----	-----
nelsonii-2	-----	-----	-----	---.	-----	-----
australe-2	-----	-----	t-----	---.	-----g	-----
bickii-2	-----	-----	-----	---.	-----	-----
bickii-3	-----	-----	-----	---.	-----	-----c--
bickii-1	-----	-----	-----	---.	-----	-----
robinsonii-1	-----	-----	-----g	---.	-----	-----
robinsonii-2	-----	-----	-----g	---.	-----	-----
sturtianum-2	-----	-----	-----	---.	-----	-----g--
sturtianum-3	-----	-----	-----	---.	-----	-----g--
sturtianum-1	-----	-----	-----	---.	-----	-----g--
londonderriense	g-----	-----	-----	---.	-c-----	-----
anapoides	g-----	-----	-----	---.	-c-----	-----
cunninghamii	g-----	-----	-----	---.	-c-g-----	-----
nobile	g-----	-----	-----	---.	-c-----	-----
marchantii	g-----	-----	-----	---.	-c-----	-----
populifolium	g-----	-----	-----	---.	-c-----	-----
costulatum	g-----	-----	-----	---a.---	-c-----	-----
enthyle	g-----	-----	-----	---.	-c-----	-----
rotundifolium	g-----	-----	-----	---.g--	-c-----	-----
pilosam	g-----	-----	-----t--	---.	-c-----	-----
exiguum	g-----	-----	-----	---.g--	-c-----	-----
pulchellum	g-----	-----	-----	---.	-c-----	-----c--
CONSENSUS	ATTGTTTATA	TCTAATAAAT	GGTTCCTTAA	TTTCGAAAAA	AAAAAAGTTA	TTGTAATTAG

	1201					1260
trilobum	-----	-g-----	-----	-----	-----	-c-----
gossypioides	-----	-g-----	-----	-----	g-----	-----
raimondii	-----	-g-----	-----	-----	-----	-c-a-----
herbaceum	-----	-----	-t-----	-----	-----	-----
arboreum	-----	-t-----	-t-----	-----	-----	-----
hirsutum	-----	-----	-t-----	-----	-----	-----
barbadense	-----	-----	-t-----	-----	-----	-----
australe-1	-----	-----	-----	-----	-----	-----
australe-3	-----	-----	-----	-----	-----	-----
nelsonii-1	-----	-----	-----	-----	-----	-----
nelsonii-2	-----	-----	-----	-----	-----	-----
australe-2	-----	-----	-----	-----	-----	-----
bickii-2	-----	-c-----	-----	-----	-----	-c-----
bickii-3	-----	-c-----	-----	-----	-----	-c-----
bickii-1	-----	-c-----	-----	-----	-----	-c-----
robinsonii-1	-----	-----	-----	-----	-----	-----t-
robinsonii-2	-----	-g-----	-----	-----	-----	-----t-
sturtianum-2	-----	-----	-----	-----	-----	-----
sturtianum-3	-----	-----	-----	-----	-----	-----
sturtianum-1	-----	-----	-----	-----	-----	-----
londonderriense	-----	-c-----	-----	-----	-----	-----
anapoides	-----	-----	-----	-----	-----	-----
cunninghamii	-----	-----	-----	-----	-----	-----
nobile	-----	-----	-----	-----	-----	-----
marchantii	-----	-a-----	-----	-----	-----	-----
populifolium	-----	-----	-----	-----	-----	-----a-
costulatum	-----	-----	-----	-----	-----	-----a-
enthyle	-----	-----	-----	-----	-----	-----
rotundifolium	-----	-----	-----	-----	-----	-----
pilosam	-----	-----	-----	-----t	-----	-----
exiguum	-----	-----	-----	-----	-----	-----
pulchellum	-----	-----	-----	-----	-----	-----
CONSENSUS	TTTGTCTCT	TGCATGTCCA	GCCAGCAAAA	TGGGTGCCGG	TGGTAGGATG	CTAGTTGACG

	1261					1320
trilobum	---t-----	-----	-----	-----	-----	-----
gossypioides	---t-----	-----	-----	-----	-----	-----
raimondii	---t-----	-----	-----	-----	-----	-----
herbaceum	-----	-a-----	-----	-----	-----	-----
arboreum	-----	-g-a-----	-----	-----	-----	-----
hirsutum	-----	-a-----	-----	-a-----	-----t-	-----
barbadense	-----	-a-----	-----	-----	-----	-----
australe-1	-----	-----	-----	-----	-----	-----
australe-3	-----	-----	-----	-----	-----	-----
nelsonii-1	-----	-----	-----	-----	-----	-----
nelsonii-2	-----	-----	-----	-----	-----	-----
australe-2	-----	-----	-----	-----	-----	-----
bickii-2	-----	-----	-----t-----	-----	-----	-----
bickii-3	-----	-----	-----	-----	-----	-----
bickii-1	-----	-----	-----	-----	-----	-----
robinsonii-1	---g---a---	-----	-----	-----	-----	-----c---
robinsonii-2	-----a---	-----	-----	-----	-----	-----c---
sturtianum-2	-----	-----	-----	-----	-----	-----c---
sturtianum-3	-----	-g-----	-----	-----	-----	-----c---
sturtianum-1	-----	-----	-----	-----	-----	-----c---
londonderriense	-----	-----	-----	-----	-----	-----
anapoides	-----	-----	-----	-----	-----	-----
cunninghamii	-----	-----	-c-----	-----	-----	-----
nobile	-----	-----	-----	-----	-----	-----
marchantii	-----	-----	-----	-----	-----	-----
populifolium	-----	-----	-----	-----	-----	-----
costulatum	-----	-----	-----	-----	-----	-----
enthyle	-----	-----	-----	-----	---t-----	-----
rotundifolium	-----	-----	-----	-----	-----	-----
pilosam	-----	-----	-----	-----	-----	-----
exiguum	-----	-----a-----	-----	-----	---t-----	-----
pulchellum	-----	-----	-----	-----	-----	-----
CONSENSUS	GTAAAAAGGA	GGAAAATCGA	GGCTCGGTCA	ATCGAGTTCC	GATCGAGAAG	CCTCCGTTTA

	1321					1375
trilobum	-----	-----	-----	-----	-----	-----
gossypioides	-----	-----	-----	-----	-----	-----
raimondii	-----	-----	-----	-----	-----	-----
herbaceum	-----	-----	-----	-----	-----	-----
arboreum	-----	-----	-----	-----	-----	-----
hirsutum	-----	-----	-----	-----	-----	-----
barbadense	-----	-----	-----	-----	-----	-----
australe-1	-----	-----	-----	-----	-----t-----	-----
australe-3	-----	-----	-----	-----	-----	-----
nelsonii-1	-----	-----	-----	-----	-----	-----
nelsonii-2	-----	-----	-----	-----c-----	-----	-----
australe-2	-----	-----	-----	-----	-----	-----
bickii-2	-----	-----	-----	-----	-----	-----
bickii-3	-----	-----	-----	-----	-----	-----
bickii-1	-----	-----	-----	-----	-----	-----
robinsonii-1	-----	-----	-----	-----	-----	-----
robinsonii-2	-----	-----	-----	-----	-----	-----
sturtianum-2	-----	-----	-----	-----	-----	-----
sturtianum-3	-----	-----	-----	-----	-----	-----
sturtianum-1	-----	-----	-----	-----	-----	-----
londonderriense	-----	-----	-----	-----	-----	-----
anapoides	-----	-----	-----	-----	-----	-----
cunninghamii	-----	-----	-----	-----	-----	-----
nobile	-----	-----	-----	-----	-----	-----
marchantii	-----	-----	-----	-----	-----	-----
populifolium	-----	-----	-----	-----	-----	-----
costulatum	-----	-----	-----	-----	-----	-----
enthyle	-----	-----	-----	-----	-----	-----
rotundifolium	-----	-----	-----	-----	-----	-----
pilosam	-----	-----	-----	-----	-----	-----
exiguum	-----	-----	-----	-----	-----	-----
pulchellum	-----	-----	-----	-----	-----	-----
CONSENSUS	CGCTCGGTCA	GATCAAGCAA	GCCATTCCGC	CCCACTGTTT	TCGCCGCTCC	CTCCT

Fig. 6.4 illustrate a pghD12-1 intron tree reconstructed using Kimura two-parameter distances (Kimura, 1980) and neighbour-joining analysis (Saitou and Nei, 1987), on the basis of the 5' UTR intron and its 3' flanking exon DNA sequences shown in Fig. 6.3. The intron sequence alone was also used to reconstruct the gene which generated an exactly same resolution (figure not shown). In general, the phylogenetic analysis of these intron sequences produced topologies consistent with taxonomic classification and previous phylogenetic studies of *Gossypium* (Endrizzi et al., 1985; Wendel and Albert, 1992; Cronn et al., 1996). However, there are also some interesting exceptional observations. Firstly, the two AD-allotetraploid cottons, *G. barbadense* and *G. hirsutum* ranked closely only with A-genome species. Secondly, all three accessions of the G-genome species, *G. bickii*, appear to be more similar to the morphologically distant C-genome species. Thirdly, one accession of *G. australe* is more similar to its sister lineage *G. nelsonii* than to other accessions of the same lineage.

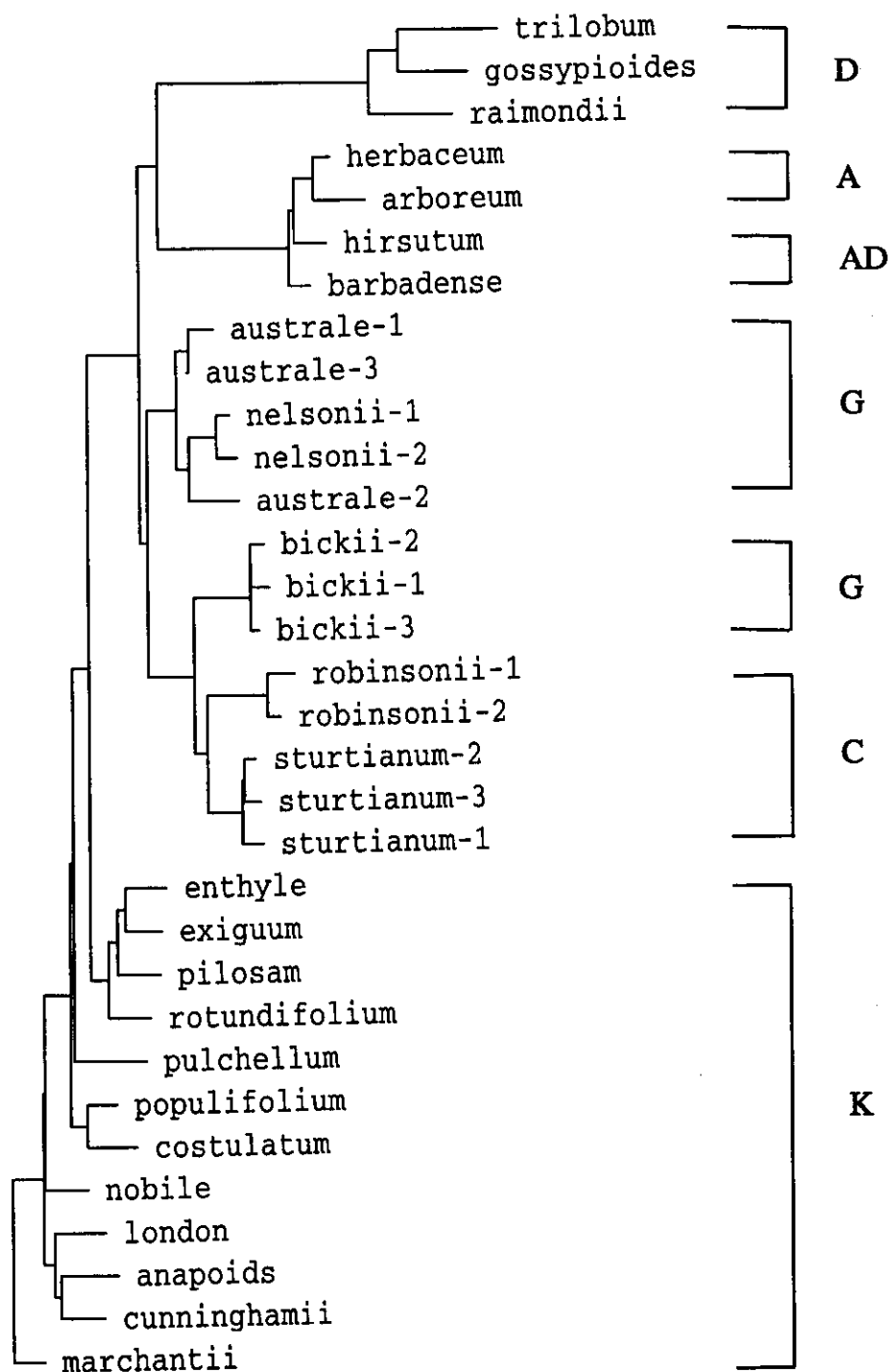


Fig. 6.4 A phylogenetic tree of 24 *Gossypium* species based on the entire 5' UTR intron and a portion of 3' flanking exon of microsomal ω -6 desaturase pghD12-1. The distance-based tree was constructed using Kimura two-parameter distances (Kimura, 1980) and neighbour-joining analysis (Saitou and Nei, 1987).

6.3.3 DNA Sequences of the Corresponding Intron from *A. thaliana* and *G. max*

The corresponding introns of the microsomal ω -6 desaturase gene FAD2 in *A. thaliana* and FAD2-1 in *G. max* have also been isolated and sequenced. Their nucleotide sequences are shown in Fig. 6.5.

A

```

1  GTCCGTCGCT TCTCTTCCAT TTCTTCTCAT TTTCGATTTT GATTCTTATT
51 TCTTTCCAGT AGCTCCTGCT CTGTGAATTT CTCCGCTCAC GATAGATCTG
101 CTTATACTCC TTACATTCAA CCTTAGATCT GGTCTCGATT CTCTGTTTCT
151 CTGTTTTTTT CTTTTGGTCG AGAATCTGAT GTTTGTTTAT GTTCTGTCAC
201 CATTAATAAT AATGAACTCT CTCATTCTATA CAATGATTAG TTTCTCTCGT
251 CTACAAAACG ATATGTTGCA TTTTCACTTT TCTTCTTTT TTCTAAGATG
301 ATTTGCTTTG ACCAATTTGT TTAGATCTTT ATTCTATTTT ATTTTCTGGT
351 GGGTTGGTGG AAATTGAAAA AAAAAAACA GCATAAATTG TTATTTGTTA
401 ATGTATTCAT TTTTGGCTA TTTGTTCTGG GTAAAAATCT GCTTCTACTA
451 TTGAATCTTT CCTGGATTTT TTAATCTTAT TGGGTTTTTA TAGTAAAAAT
501 ACATAATAAA AGGAAAACAA AAGTTTATA GATTCTCTTA AACCCTTAC
551 GATAAAAGTT GGAATCAAAA TAATTCAGGA TCAGATGCTC TTTGATTGAT
601 TCAGATGCGA TTACAGTTGC ATGGCAAATT TTCTAGATCC GTCGTCACAT
651 TTTATTTTCT GTTTAAATAT CTAAATCTGA TATATGATGT CGACAAATTC
701 TGGTGGCTTA TACATCACTT CAACTGTTTT CTTTGGCTT TGTTTGTCAA
751 CTGGTTTTTC AATACGATTT GTGATTTTCA TCGCTGAATT TTAATACAA
801 GCAAACGATG GTTAACCACA AGCAAGAGAT GTGACCTGCC TTATTAACAT
851 CGTATTACTT ACTACTAGTC GTATTCTCAA CGCAATCGTT TTTGTATTTT
901 TCACATTATG CCGCTTCTCT ACTCTTTATT CCTTTTGGTC CACGCATTTT
951 CTATTTGTGG CAATCCCTTT CACAACCTGA TTTCCCACTT TGGATCATTT
1001 GTCTGAAGAC TCTCTTGAAT CGTTACCACT TGTTTCTTGT GCATGCTCTG
1051 TTTTTTAGAA TTAATGATAA AACTATTCCA TAGTCTTGAG TTTTCAGCTT
1101 GTTGATTCTT TTGCTTTTGG TTTTCTGCAG

```

B

```

1  GTAATTAATA TTGTGCCTGC ACCTCGGGAT ATTTTCATGTG GGGTTCATCA
51 TATTTGTTGA GGAAAAGAAA CTCCCGAAAT TGAATTATGC ATTTATATAT
101 CCTTTTTCAT TTCTAGATTT CCTGAAGGCT TAGGTGTAGG CACCTAGCTA
151 GTAGCTACAA TATCAGCACT TCTCTCTATT GATAAACAAAT TGGCTGTAAT
201 GCCGCAGTAG AAGACGATCA CAACATTTTCG TGCTGGTTAC TTTTGTGTTT
251 ATGGTCATGA TTTCACTCTC TCTAATCTCT CCATTCAATT TGTAGTTGTC
301 ATTATCTTTA GATTTTTCAC TACCTGGTTT AAAATTGAAG GATTGTAGTT
351 CTGTTGGTAC ATATTACACA TTCAGCAAAA CAACTGAAAC TCAACTAAAC
401 TTGTTTATAC TTTGACACAG

```

Fig. 6.5 Nucleotide sequences of the 5' UTR intron of FAD2 gene in *A. thaliana* (A) and intron 1 of FAD2-1 in *G. max*. (B). The 5' and 3' exon/intron boundaries :GT/AG: are in bold type.

A striking similarity exists between *A. thaliana* and *G. hirsutum* in terms of intron length and location. The *A. thaliana* intron is 1130 bp long, located at 4 bp upstream the translation initiation site. However, the intron 1 of *G. max* is 320 bp, much shorter than its counterparts in *A. thaliana* and *G. hirsutum*. The intron 1 from *G. max* is located within the coding region, just behind the translation initiation site ATG. Therefore, the *G. max* intron is not an 5' UTR intron, although it is present at the beginning of the coding region.

6.3.4 PCR Amplification of D-subgenome Intron from Tetraploid *Gossypium* Species

The primer *d12DS1* was designed based on the D-genome-specific fragment of the pghD12-1 intron from D-genome diploids. When it was used in a PCR experiment to amplify the pghD12-1 intron from allotetraploids, it was expected that only the intron fragment of D-subgenome origin would be obtained. In the PCR experiment illustrated in Fig. 6.6A, A-genome diploid species *G. herbaceum* and *G. arboreum* were used as negative controls and D-genome species *G. raimondii*, *G. trilobum* were used as positive controls. As shown in Fig. 6.6A, except for the negative control (lane 3-4), a distinct band of expected size (640 bp) was observed in each of all other four species including the two allotetraploids, *G. barbadense* (lane 1) and *G. hirsutum* (lane 2) and two positive controls, *G. raimondii* (lane 5) and *G. trilobum* (lane 6).

When a new primer *d12it-1* which is located at about 130 bp downstream of 5' exon/intron boundary, but about 570 bp upstream the *d12DS1* primer was used in combination with the original *d12A4* primer in a PCR with genomic DNA from *G. barbadense* and *G. hirsutum*, a PCR product of the expected size was obtained in each of the two tetraploids (lane 1 and 2 in Fig. 6.6B). Further investigation by PCR using the above PCR product as DNA template and the *d12A4* primer in combination with the D-genome-specific primer *d12DS1* as a nested primer demonstrated that only the *G. barbadense* derived PCR product contains both A- and D-subgenome introns (lane 3 in

Fig. 6.6B), whereas the *G. hirsutum* product does not contain the D-subgenome intron (lane 4 in Fig. 6.6B).

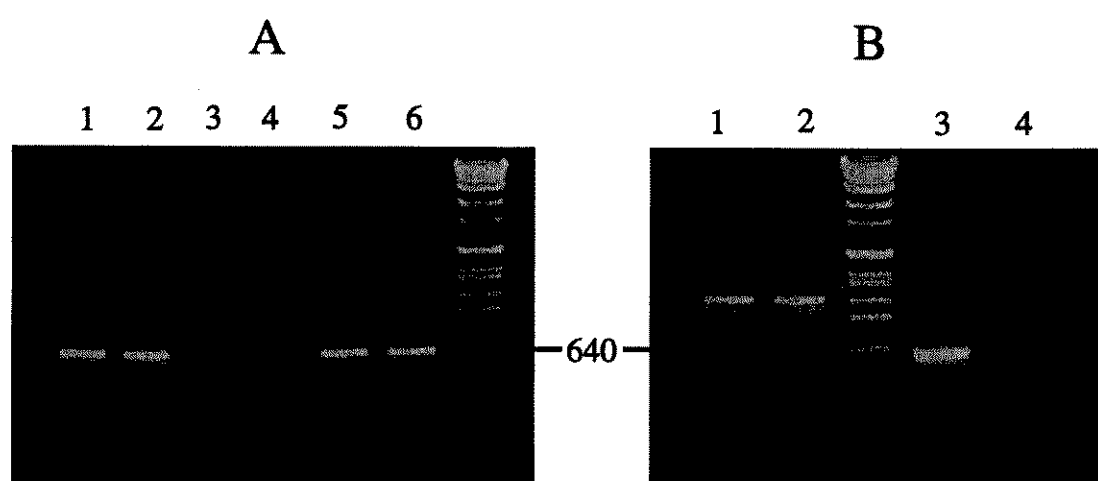


Fig. 6.6 PCR amplification of the pghD12-1 intron of D-subgenome origin from allotetraploids.

(A) PCR using D-genome-specific primer *d12DS1* and *d12A4*. The size of the PCR product is indicated on the left. The genomic DNA templates are from: *G. barbadense* (lane 1), *G. hirsutum* (lane 2), *G. herbaceum* (lane 3), *G. arboreum* (lane 4), *G. raimondii* (lane 5) and *G. trilobum* (lane 6). The last lane is a SPP-1/EcoRI DNA ladder (Bresatec).

(B) PCR using *d12it-1* and *d12A4* primers. Lane 1: *G. barbadense*; lane 2: *G. hirsutum*; lane 3: SPP-1/EcoRI DNA ladder. Second round of PCR using *d12DS1* as a nested primer in combination with *d12A4* and the PCR product from first round of PCR (lane 1 or 2) as DNA templates. Lane 3: using lane 1 product as template; Lane 4: using lane 2 products as template.

On the basis of the above PCR results, the D-subgenome intron fragment of *G. hirsutum* (lane2 in Fig. 6.6A) was cloned and sequenced. The PCR products shown lane1 in Fig. 6.6B contains fragments of introns of both A- and D-subgenome origin in *G. barbadense*. Subsequently, the D-subgenome intron was selected, cloned and sequenced. Fig. 6.7A illustrates the DNA sequence comparison of both A- and D-subgenome intron sequences from the two allotetraploid cottons (*G. barbadense* and *G. hirsutum*) and those of the close relatives of their putative progenitors. The conservation of nucleotide substitutions between the D-subgenome sequences from tetraploids and the D-genome diploid sequences, especially *G. raimondii*, the closest relative of the D-subgenome donor, is apparent. The nucleotide substitution rate in *G. barbadense* D-subgenome intron is not obviously higher than the D-genome sequences. Since the *dl2it-1* primer was successfully used to amplify the D-subgenome intron from *G. barbadense* but failed from *G. hirsutum*, it is reasonable to predict some sequence variations between the D-subgenome introns from *G. barbadense* and *G. hirsutum*. However, only a short and relatively conserved 3' portion of the D-subgenome intron has been isolated from *G. hirsutum*, the sequence divergence between these two D-subgenome introns is not clear. When the DNA sequences shown in Fig. 6.7A were used to reconstruct a distances-based gene tree, the identity of the partial D-subgenome intron sequences are clearly demonstrated (Fig. 6.7B).

Fig. 6.7 The subgenomic intron sequences of pghD12-1 gene from allotetraploids and those closely related to the putative diploid progenitors.

(A). DNA sequence alignment. A consensus was drawn using GCG program PRETTY. The intron starts with GT and ends with AG, which are in bold case.

(B). A distances-based phylogenetic tree.

(A)

	1					60
herbaceum	-----	-----	-----	-----	-----	-----
arboreum	-----	-----t	-----	-----	-----	-----
hirsutum (A)	-----	-----	-----	-----	-----	-----
barbadense (A)	-----	-----	-----	-----	-----	-----
gossypioides	-----	-----	-----	-----	-----c-----	-----
raimondii	-----	-----	-----	-----	-----c-----	-----
trilobum	-----	-----	-----	-----	g-----c-----	-----
CONSENSUS	GTACATTCT	CTTTAATTC	CTTCATTTT	TTTCCATTTC	ATGTTTCTCT	TGTTAATTTT
	61					120
herbaceum	-----	-----	-----	-----	-----	-----
arboreum	-----	-----	-----	-----	-----	-----
hirsutum (A)	-----	-----	-----	-----	-----g-----	-----
barbadense (A)	-----	-----	-----	-----	-----	-----
barabadense (D)						
gossypioides	-----	-----	-----g	-----	-----	-----
raimondii	-----	-----	-----	-----	-----	-----
trilobum	-----	-----	-----	-----	-----c-----	-----
CONSENSUS	GCATTGAAGT	GATAAATTG	AGTGAATGAT	GTTTGGTATA	TCTTCTTAGT	AACTGACCTT
	121					180
herbaceum	-----	-----	-----	-----	-----	-----
arboreum	-----	-----	-----	-----	-----	-----
hirsutum (A)	-----	-----c	-----c	-----	-----g-----	-----
barbadense (A)	-----	-----	-----	-----	-----	-----
barabadense (D)	-----	-----	-----	-----	-----t-t	c---tga---
gossypioides	-----	-----	-----	-----t-t	-----tga---	-----gt---
raimondii	-----	-----	-----	-----	-----t-t	c---tga---
trilobum	-----	-----	-----	-----t-t	-----tga---	-----t---
CONSENSUS	TTGAAAATAC	TAGCATTTT	TTTTTTTAAT	ATCAAGTGAA	AGAAGAAGAA	GAAGAAGAAG
	181					240
herbaceum	-----	-----	-----	-----	-----c-----	-----g---
arboreum	-----g---	-----	-----	-----	-----	-----g---
hirsutum (A)	-----	-----	-----	-----t	-----	-----
barbadense (A)	-----	-----	-----	-----	-----	-----
barabadense (D)	-----	-----	-----	-----	-----c-----	-c-tagatc-
gossypioides	-----	-----	-----	-----	-----c-----c	-----
raimondii	-----	-----	-----	-----	-----c-----c	-----
trilobum	-----t-	-----	ag-----	-----	-----c-----c	-----
CONSENSUS	AAGAAGAAGA	ATTTCGCCAT	GCAAAAAGCT	TTTTAAGGC-	TTTTTTTTTT	CTTAGATCAA
	241					300
herbaceum	-----	-----	-----	-----	-----	-----
arboreum	-----c-	-----	-----	-----	-----	-----
hirsutum (A)	-----	-----	-----	-----	-----	-----
barbadense (A)	-----	-----	-----	-----g-----	-----	-----
barabadense (D)	--a--ta--t	g--tac-ta-	-----	-----c-----	-----	-----
gossypioides	-----	-----	-----	-----	-----	-----
raimondii	-----	-----	-----	-----	-----	-----
trilobum	-----	-----	-----	-----	-----	-----
CONSENSUS	AATTTATTTG	TTTACTTATA	CTGTTCTTTT	AAGCCTGAAG	AAAGAAGCCA	TGGTTTCAAT
	301					360
herbaceum	-----	-----g-----	-----	-----	-----a-----	-----
arboreum	-----	-----	-----	-----	-----	-----
hirsutum (A)	-----	-----	-----	-----	-----	-----
barbadense (A)	-----	-----	-----	-----	-----	-----
barabadense (D)	-----	-----	-----a-----	-----	-----	-----
gossypioides	-----	-----	-----	-----c-----	-----	-----
raimondii	-----	-----	-----a-----	-----	-----	-----
trilobum	-----	-----	-----t	-----	-----	-----
CONSENSUS	TTTTGAGAGT	TTTATAAATC	CCAGATACCA	GAGAGCTTCA	TCGTTTATTC	ATATATTTTT

	361				420
herbaceum	g-----	-----	-----	-t-----	-----c--
arboreum	g-----	-----	-----	-t-----	-----c--
hirsutum (A)	g-----	-----	-----	-t-----	-----c--
barbadense (A)	g-----	-----	-----	-----	-----c--
barabadense (D)	a-----	-----	-----	-----aa-	-----t--t
gossypioides	a-----	-----	-----	-----	t-----t--
raimondii	a-----	-----	-----	-----	-----t--
trilobum	a-----	-----	-----	-----	-----t--
CONSENSUS	-AACATTTT	TAAAGCAAGA	ACTTGTGATT	TGTTTTTGGT	AAAATATGCA ATAAAT-TTC

	421				480
herbaceum	-----	-----	-----c--	---g---t	---a--t--- ---t---aaa
arboreum	-----	-----	-----c--	---g---t	---a--t--- ---t---aaa
hirsutum (A)	-----	-----	-----c--	---g---t	---a--t--- ---t---aaa
barbadense (A)	-----	-----	-----c--	---g---t	---a--t--- ---t---aaa
barabadense (D)	-----	-----	-----tt-	-c--c---a	-a-t--a--a aag-aa-ttt
gossypioides	-----	-----	-----t-	---c---a	ga-t--a--- t---actttt
raimondii	-----	-----	---g--t-	---t--a-	t---actttt
trilobum	-----	-----	-----t-	---c---a	ga-t--a-g- t---actttt
CONSENSUS	ATATTTTTCG	TAAATTTTAA	ATTTAAT-CT	TTTA-TTTT-	AGA-TT-AAT CTTT-TA---

	481				540
herbaceum	-----	-----	-----	---g---t	-a--t--- ---t---aaa
arboreum	-----	-----	-----	---a---t	-a--t--- ---t---aaa
hirsutum (A)	-----	-----	-----	---a---t	-a--t--- ---t---aaa
barbadense (A)	-----	-----	-----	---a---t	-ag-t--- ---t---aaa
barabadense (D)	-aa-at....			---acc---ca	-t--a---a
gossypioides	-tagat--t-	-aa-a--ccg	t-aaatttta	aca-tg-t-	-a-ta--ta -t--a---a
raimondii	-t--ga--tt	-aa.....		---acg---ca	-t--a---a
trilobum	-t--ga--tt	-aa.....		---acg---ta	-t--a---a
CONSENSUS	TCTACGTTAA	ATTTTAA-C	AT-----	---TATTAAA	ATT-TTTTA- T-AA-TTAGA

	541				600
herbaceum	-----t--	-----	-----	-c---t--	-c-t-----
arboreum	-----t--	-----	-----	-c---t--	-c-t-----
hirsutum (A)	-----t--	-----	-----	-c---t--	-c-t-----
barbadense (A)	-----t--	-----	-----	-c---t--	-c-t-----
barabadense (D)	-----a--	-----	-----a-	-tt--a--	caca-----
gossypioides	-----a--	-----	-----tg-	-tt--a--	cata-----
raimondii	-----a--	-----	-----a-	-tt--a--	caca-----
trilobum	-----a--	-----a-	-----ta-	-tc--a--	taca-----
CONSENSUS	TTATTAT-AG	TAATTGAGTA	TTTTTTTATT	T-AAAA-TTT	A-A-CATCAA ATTTAAAAAA

	601				660
herbaceum	-----	-----	-----	-----	-----
arboreum	-----	-----	-----	-----	-----
hirsutum (A)	-----	-----	-----	-----	-----
barbadense (A)	-----	-----	-----	-----	-----
barabadense (D)	-----	-----	-----	-tg-a-ag-	-----t--
gossypioides	-----	-----	-----	-tg-a-ag-	-----c--
raimondii	-----	-----	-----	-----	-----c--
trilobum	-----	-----t--	-----	-tg-g-ag-	-----c--
CONSENSUS	AAAAAACAAA	GTAAACACTT	GCACTTGATT	TCAATATCTA	AAAGGA-TAA ATTT-AAATT

	661				720
herbaceum	...--g--	a-g--t--	-----	-----	-g--t--- ---t---ta
arboreum	...--g--	a-g--t--	-a-----	-g-----	-g--t--- ---t---ta
hirsutum (A)	...--g--	a-g--t--	-----	-----	-g--t--- ---t---ta
barbadense (A)	...--g--	a-g--t--	-----	-----	-g--t--- ---t---ta
barabadense (D)	---g--a--	g-a--a--	-c-----	-----	-a--c--c-- ---a---ag
gossypioides	---g--a--	g-a--a--	-----	-----c	-a--c--c-- ---a---ag
raimondii	---t-g--a--	g-at-a--	-c-----	-----	-a--c--c-- ---a---ag
trilobum	-----a--	g-a--a--	-c-----	-----	-a--c--c-- ---a---ag
CONSENSUS	TTCAATA-AA	-G-CT-AATT	TTAAATTTTC	AAAGAGTATA	G-GA-TTCTC TAC-TTTT--

	721					780
herbaceum	g-----gt-		..a-----	-----	--a-t-ac-	-----
arboreum	g-----gt-		..a-----	-----	--a-t-ac-	-----
hirsutum (A)	g-----gt-		..a-----	-----	--a-t-ac-	-----
barbadense (A)	g-----gt-		..a-----	-----	--a-t-ac-	-----
hirsutum (D)			...	-----	--t-----	-----
barabadense (D)	at-----aa-	-----	--g-----	-----	--c-----	-----
gossypioides	at-----aa-	-----g-	--g-----	-----	--t-----	-----
raimondii	a-----aa-	-----	--g-----	-----	--t-----	-----
trilobum	at-----aa-	-----	--g-----	-----	--t--g-	-----
CONSENSUS	-ATTTAA--T	TTAAATCTAA	CA-TTAACAC	TTCTTTAATT	AC-TTACGAT	AAATTTAACT

	781					840
herbaceum	.-----g-	-----	-----	----ag--	---c----	c-----t-
arboreum	.-----g-	-----	-----	----ag--	---c----	c-----t-
hirsutum (A)	.-----	-----	-----	-----g-	---c----	c-----t-
barbadense (A)	.-----	-----	-----	-----g-	---c....	
hirsutum (D)	-----	-----g	-----	-----	-----	-----
barabadense (D)	-----	-----g	-----	-a-----	-----	-----
gossypioides	-----	-----	-----	-a-----	c--g--g-	-----
raimondii	-----	-----g	-----	-a-----	-----g-	-----
trilobum				...	-a-----g-	-----
CONSENSUS	AAAAAATTAC	AATATTAATA	GTTAAATTA	AATTTTGAAA	AGTATAAAGA	TTAAATGTGA

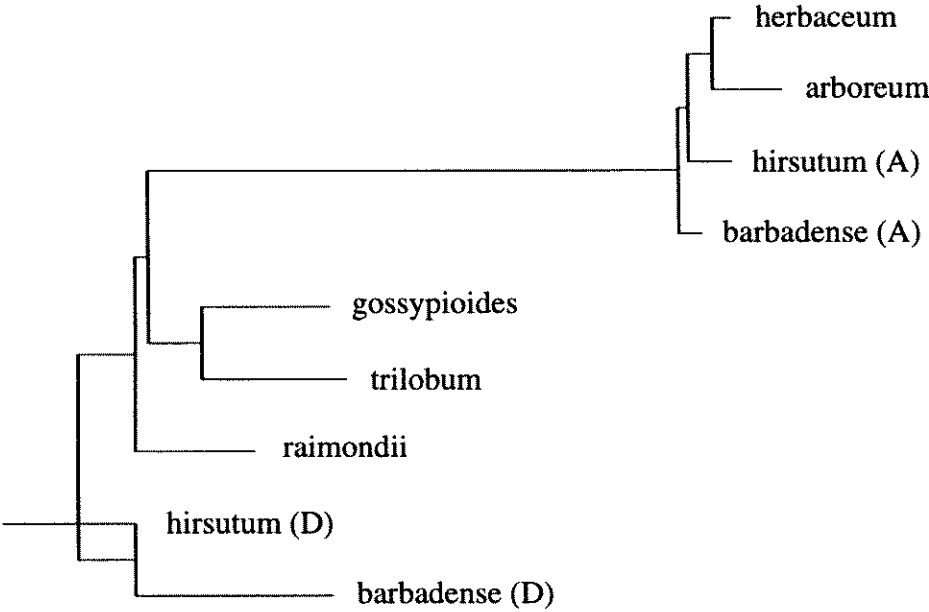
	841					900
herbaceum	-----	--a-----	-----	-----	-----	-c---a--
arboreum	-----	--a-----	-----	-----	-----g--	-c---a--
hirsutum (A)	-----	--a-----	-----	-----	-----	-c---a--
barbadense (A)	-----	--a-----	-----	-t-----	-----	-c---a--
hirsutum (D)	-----	--c-----	-----	t-----	-----	-----g
barabadense (D)	-----	--c-----	-----	t-----	-----	-----g
gossypioides	-----	-----	-----	t-----	-----	-----
raimondii	----g----	-----	-----	t-----	-----	----c---
trilobum	-----	-----	-----	-----	-----	-----
CONSENSUS	AATTTTCAAA	AAGTATAGGA	AGTTATAGTA	CATTTTAACC	TTTATTTATT	TTATATCTGA

	901					960
herbaceum	-----	-----	---a-----	-----	---g-----	-----
arboreum	-----	-----	---a-----	-----	---g-----	-a-----
hirsutum (A)	-----	-----	---a-----	-----	---g-----	-----
barbadense (A)	-----	-----	---a-----	-----	---g-----	-a-----
hirsutum (D)	-----	-----	-----	-----	-----	-----
barabadense (D)	---a-----	-----	-----	-----	-----	-----
gossypioides	-----	-----	-----	-----	--t-----	-----
raimondii	---a-----	-----	-----	-----	---	-----
trilobum	-----	-----	-----	--c-----	-----	-----
CONSENSUS	TGAGGTTTCCT	GCATGCACCG	AAGATGTCAC	CTTTTGCCAG	TA-TTTTCCA	GTGGCTTGTT

	961					1020
herbaceum	-----	-----	-----	-----	--a-----	-----c
arboreum	-----	-----	-----	-----	--a-----	-----c
hirsutum (A)	-t-t-----	-----	-----	-----	--a-----	-----c
barbadense (A)	-----	-----	-----	-----	--a-----	-----c
hirsutum (D)	-----	-----	-----	-----	--t-----	-----
barabadense (D)	-----	-----	-----	-----	--t-----	-----
gossypioides	-----	-----	-----	-----	--t-----	-----
raimondii	--g-----	-----	-----	-----	--c-----	-----
trilobum	-t-t-----	-----	-----	-----	--t-----	-----
CONSENSUS	TCTCTCAAAA	CTACCTTGAA	TCTTGAGACA	GAATTAAATA	TA-TTTTGGC	CTTTTCTTCA

	1021					1080
herbaceum	---.,-----	-----c-	-----	-----	-----	-----
arboreum	---.,-----	-----c-	-----	--a-----	-----	-----
hirsutum (A)	---.,-----	-----	-----	-----	-----	-----
barbadense (A)	---.,-----	-----	-----	-----	-----	-----
hirsutum (D)	t---.,-c---	-----	-----	-----	-----	-----
barabadense (D)	t--tc-c---	-----	-----	-----	-----	-----
gossypioides	---.,-----	-----	-----	-----	-----	-----
raimondii	t---.,-----	-----	-----t---	-----	-----	-----
trilobum	---.,-----	-----	-----g-	-----	-----	-----
CONSENSUS	ATT--TTTCT	CTCTCTATTT	TCTTTTAAAA	ATTGCTTTAG	AGAATTCAGA	AAAAATACTT
	1081					1140
herbaceum	-----	-----g---	-----g-	-----	-----	-----
arboreum	-----	-----g-c-	-----g-	-----	-----	-----
hirsutum (A)	-----	-----g---	-----g-	-----	-----	-----
barbadense (A)	-----	-----g---	-----g-	-----	-----	-----
hirsutum (D)	-----	-----	-----	-----	-----	-----
barabadense (D)	-----	-----	-----	-----	-----	-----
gossypioides	-----	-----	-----	-----	-----	-----
raimondii	-----	-----	-----	-----	-----	-----
trilobum	-----	-----	-----	-----	-----	-----
CONSENSUS	TCCAACACGA	AAATTTCTTC	AAATTTATTG	TTTATATCTA	ATAAATGGTT	GCTTAATTTT
	1141					1186
herbaceum	aa----a---	-----	a-----	-----a-	-----	-----
arboreum	aa----a---	-----	a-----	---t---a-	-----	-----
hirsutum (A)	aa----a---	-----	a-----	-----a-	-----	-----
barbadense (A)	aa----a---	-----	a-----	-----a-	-----	-----
hirsutum (D)	-----	-----	-----	-----	-----	-----
barabadense (D)	-----	-----	-----	-----	-----	-----
gossypioides	-----	-----	-----	-----	-----	-----
raimondii	-----	-----	-----	-----	-----	-----
trilobum	-----	-----	-----	-----	-----	-----
CONSENSUS	GGAAAACAAA	AGTTATTGTA	GTTAGTTTGTG	CTTCTTGCGT	GTCCAG	

(B)



6.3.5 Southern Blot Analysis at High-Stringency Conditions

A Southern blot with EcoRI- and HindIII-digested genomic DNAs from *G. barbadense*, *G. hirsutum* cvs Deltapine and Siokra, *G. herbaceum*, *G. raimondii* and *G. robinsonii* was probed by a DNA fragment corresponding to the entire 3' UTR of pghD12-1. This is similar to the previously described experiment in Chapter Five, except that the blot was washed under a higher stringency conditions following hybridisation. As shown in Fig. 6.8, in both EcoRI and HindIII restriction fragment patterns, there are two distinct hybridising bands in the allotetraploid cottons (lane 1-3), while a single band in each of the diploid species (lane 4-6). Based on the size similarities with the band of the A- and D-genome diploids, the two bands in each allotetraploid cotton are assumed to represent their respective subgenomic origins of the pghD12-1 gene. In both EcoRI and HindIII-restriction digestions of allotetraploids, the top band corresponds the D-subgenome, while the lower band corresponds the A-subgenome. A very important observation of Fig. 6.8 is that the signal of D-subgenome band in both EcoRI and HindIII patterns is obviously less intense than the A-subgenome band of allotetraploids. This clearly indicates the sequence divergence of the 3' UTR of D-subgenome pghD12-1 gene in allotetraploid cottons.

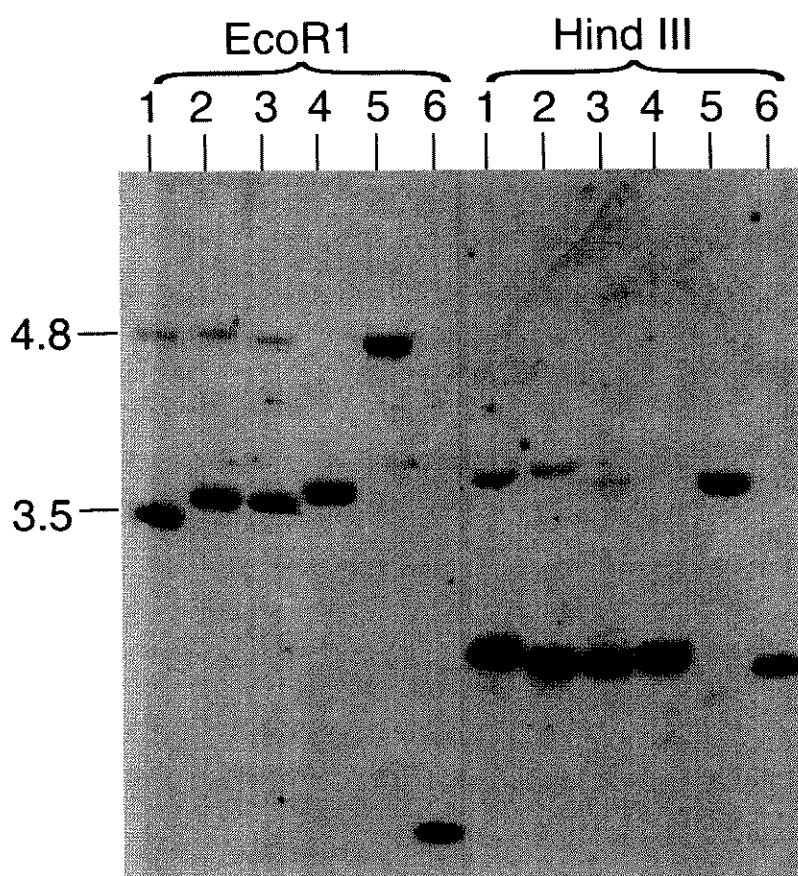


Fig. 6.8 Southern blot analysis of EcoRI- and HindIII-digested genomic DNAs from *G. barbadense* (lane 1), *G. hirsutum* cv Deltapine-16 (lane 2), cv Siokra (lane 3); *G. herbaceum* (lane 4); *G. raimondii* (lane 5) and *G. robinsonii* (lane 6). The DNA probe was a DNA fragment corresponding to the entire 3' UTR of pghD12-1. The blot was washed at a high stringency condition.

6.3.6 Construction of a 35S-Intron-GUS Fusion System

In this experiment, the full length pghD12-1 intron and two short flanking exon sequences from *G. hirsutum* were inserted into the pBI121 vector, between CaMV35S promoter and the GUS gene in a sense orientation (Fig. 6.9). The design of primers allowed the maintenance of the original splicing site at both 5' and 3' exon/intron boundaries. The first ATG at the 3' flanking exon region of insert sequence was mutated so that the normal translation of the GUS reporter gene will not be affected.

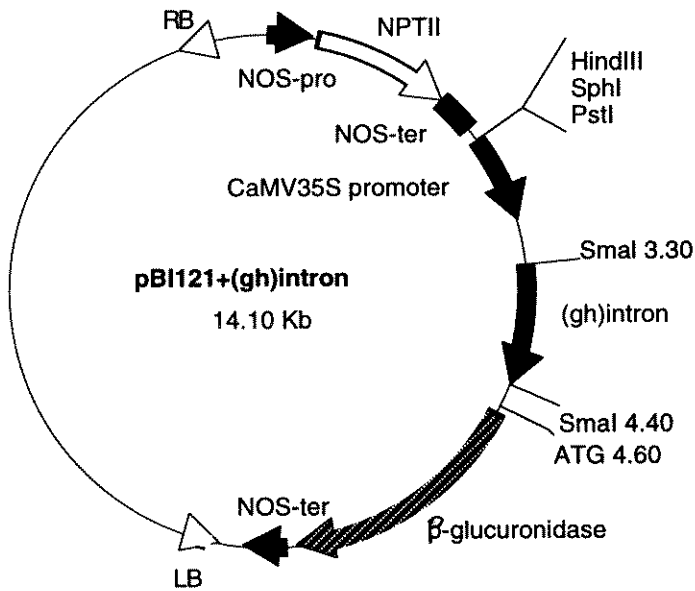


Fig. 6.9 A 35S-intron-GUS chimeric construct. The pghD12-1 intron was inserted into a SmaI site between 35S promoter and GUS gene in a sense orientation.

6.4 DISCUSSION

6.4.1 Sequence Characteristics of the Microsomal ω -6 Fatty Acid Desaturase Intron

It has been previously demonstrated that there is only one intron in *A. thaliana* FAD2 gene and it is located at the 5' UTR of the gene (Okuley et al., 1994). Although the genomic clone of the microsomal ω -6 desaturase gene has not been isolated in the current study, PCR experiments with *G. hirsutum* genomic DNA, and various combinations of primers covering the entire region of the cDNA pghD12-1, indicated that there was no other intron besides the 5' UTR intron in the pghD12-1 gene in *G. hirsutum*. However, it is not known whether this is also the case for FAD2-1 gene in *G. max*. Therefore, the *G. max* intron is referred to as intron 1 in this study. Table 6.2 summaries some of the characteristic features of the microsomal ω -6 desaturase intron from *G. hirsutum*, *G. max* and *A. thaliana*.

Table 6.2 A Summary of the Features of the Microsomal ω -6 Desaturase Intron from *G. hirsutum*, *G. max* and *A. thaliana*

FEATURES	<i>G. hirsutum</i>	<i>G. max</i>	<i>A. thaliana</i>
intron position*	-9	+4	-5
intron length	1133	420	1130
A content (%)	34.2	27.9	24.2
T content (%)	42.2	37.1	42.7
C content (%)	11.9	18.3	18.5
G content (%)	11.7	16.7	14.6
5' exon/intron boundary	AG:GTACAT	TG:GTAAAT	AG:GTCCGT
3' intron/exon boundary	TCCAG:C	CACAG:G	TGCAG:A
branchpoint	TTAGT	-	TTGAT
branchpoint position**	-23	-	-26

*relative to the translation start site (1st ATG)

**relative to the 3' intron/exon boundary and the last nucleotide of intron, G, is numbered as -1.

Higher plant pre-mRNA introns vary widely in length, but around two thirds are shorter than 150 bp and most fall within a size range of 80-139 bp (Simpson et al., 1993; Filipowicz et al., 1995). There are very few plant introns are over 1 kb. Hence, the 5' UTR intron of the *pghD12-1* gene is a relatively large intron. It is also interesting that the length and location of the intron are astonishingly conserved between *pghD12-1* in *G. hirsutum* and *FAD2* in *A. thaliana*, whereas the *G. max* *FAD2-1* intron is only less than one third of their sizes and it is located within the coding region.

All three introns follow the role of the 5' and 3' splice site binucleotides consensus (:GT/AG:). In comparison with the wider 5' splice site consensus in higher plants (AG/GTAAGT) (Simpson and Filipowicz, 1996), six out of eight nucleotides are conserved in each of the three introns although the variable bases are different in each intron.

Similar to other higher plant introns, in particular dicot plant introns, the AT contents of these introns are obviously higher than their respective flanking exons. In both *A. thaliana* and *G. max*, AT content is about twice that of the GC content, while the *G. hirsutum* intron AT content is about three times that of the GC content. In terms of individual base content, G and C are similar. However, T content is obviously higher than A. When compared to flanking exon sequences, the T content of these introns increases markedly, but the content of A changes little. Therefore references to the T-richness may be more appropriate. It is also observed that although AT-rich sequence is spread throughout the entire length of introns, extended stretches of AT-rich sequences also exist (Fig. 6.3 and 6.5A,B). There is a small enrichment of T residues close to the 3' splice site in all three introns. It has been believed that the T-rich sequence upstream the 3' splice site can influence splice site selection in plant introns (Lou et al., 1993; Lüehrsen and Walbot, 1994).

Recently, the first precisely mapped plant intron branch point has been reported to be UUGAU at 32 bp upstream the 3' splice site in the Rubisoc activase gene of *A. thaliana* (Liu and Filipowicz, 1996). In the current experiment, the branch point of *G. hirsutum*

intron was estimated to be in the position of -23 from 3' splice site and -26 for *A. thaliana*. However, the branch point of soybean is not so obvious and not be estimated.

Attempts to align these three intron sequences have failed to reveal significant sequence similarities. There is also no obvious conserved element among them. The lack of sequence similarity among these three dicots may indicate that the microsomal ω -6 desaturase intron is not phylogenetically informative at the high taxonomic level. However, this remains to be proved as more DNA sequences of the corresponding intron from plants at various taxonomic levels will be available in the future.

6.4.2 The Sequence Variations in the N-terminal of pghD12-1 ORF among *Gossypium* Species

The 5' UTR intron of pghD12-1 gene and its flanking exon sequences were amplified from all the 32 accessions of the 24 *Gossypium* species examined. The nucleotides and their deduced amino acid sequences in the 3' flanking exon regions were first examined for their phylogenetic information. Fig. 6.10 illustrate the sequence alignment of the first 48 amino acid residues in the N-terminal portion of pghD12-1 gene in these *Gossypium* species. There are some apparent conservations of amino acid substitutions within some genomic groups. For example, the two A-genome diploid species and the A-subgenome in allotetraploids share an unique lysine (K) substitution at position 16, while the three D-genome diploid species and the D-subgenome from allotetraploids share a proline (P, with the exception of *G. gossypioides*) and an isoleucine (I) at position 8 and 12, respectively. The three *G. bickii* accessions also share this proline (P) at position 8. The liaison between *G. bickii* and D-genome species on this particular substitution is incongruent with previous phylogenetic analysis. It is probably just coincidental. There are also some random amino acid substitutions in seven accessions, while the rest 16 accessions share a perfect match, without variations at the amino acid level (Fig. 6.10). The phylogenetic analysis on the basis of the 3' flanking exon sequence alone failed to generate a fully resolved topology (not shown). This is

probably because of its high sequence homology and short length which place practical limitations on its ability to record significant evolutionary history.

	1				48
raimondii	-----p--	-i-----	-----	-----	-----
hirsutum (D)	-----p--	-i-----	-----	-----	-----
barbadense (D)	-----p--	-i-----	-----	-----	-----
trilobum	-----p--	-i-----	-----	-----	-----
gossypioides	-----	-i-----	-----	-----	-----
herbaceum	-----	-----k--	-----	-----	-----
arboreum	-----	-----k--	-----	-----	-----
barbadense (A)	-----	-----k--	-----	-----	-----
hirsutum (A)	-----	-----k--	-----	-----	-----
robinsonii-2	-----	-----	-----	-----	-----
australe-1	-----	-----	-----	-----	-----
australe-3	-----	-----	-----	-----	-----
nelsonii-1	-----	-----	-----	-----	-----
australe-2	-----	-----	-----	-----	-----
sturtianum-1	-----	-----	-----	-----	-----
sturtianum-2	-----	-----	-----	-----	-----
londonderriense	-----	-----	-----	-----	-----
anapoids	-----	-----	-----	-----	-----
nobile	-----	-----	-----	-----	-----
marchantii	-----	-----	-----	-----	-----
enthyle	-----	-----	-----	-----	-----
rotundifolium	-----	-----	-----	-----	-----
pulchellum	-----	-----	-----	-----	-----
populifolium	-----	-----	-----	-----	-----
costulatum	-----	-----	-----	-----	-----
robinsonii-1	-----	-r-----	-----	-----	-----
cunninghamii	-----	-----a--	-----	-----	-----
exiguum	-----	-----q--	-----	-----	-----
sturtianum-3	-----	-----s--	-----	-----	-----
pilosam	---v-----	-----	-----	-----	-----
bickii-3	-----p--	-----	-----	-----	-----
bickii-1	-----p--	-----	-----	-----	-----
bickii-2	-----p--	-----l-	-----	-----	-----
nelsonii-2	-----	-----	-----x--	-----s--	-----
Consensus	MGAGGRMLVD	GKKEENRGSV	NRVPIEKPPF	TLGQIKQAIP	PHCFRRSL

Fig. 6.10 An alignment of the deduced amino acid sequences in the N-terminal portion (48 amino acid residues) of the pghD12-1 gene from 32 accessions of 24 *Gossypium* species. A consensus sequence is drawn by using a GCG program PRETTY. Those amino acid residues congruent with consensus are indicated by a "-" and amino acid substitutions are also shown.

6.4.3 The Sequence Composition of a Microsatellite Within the pghD12-1 Intron

A microsatellite type fragment with various numbers of GAA trinucleotide repeats in the pghD12-1 intron sequences from all the *Gossypium* species examined (Table 6.3). The number of continuous (GAA) repeats ranges from 4 repeat in *G. raimondii* or D-subgenome *G. barbadense* to 14 in *G. populifolium*. Because of the great variability in the number of repeat units at most microsatellite loci and relative ease of scoring, microsatellites are now generally considered as one of the most powerful genetic markers (Tautz, 1989; Kashi et al., 1990). The majority of mutations at microsatellite loci are stepwise in nature, changing allelic sizes by one or a very small number of repeats. Methods of genetic distances has been designed specifically to apply to microsatellites generally assuming Ohta and Kimura's (1973) stepwise mutation model (SMM) or one of its generalisations (Levinson and Gutman, 1987; Schlötterer and Tautz, 1992).

However, in contrast to the classical microsatellite, sequence diversity in the microsatellite region in the pghD12-1 intron is not solely the result of number variation in the primary tandem repeat unit. Interspecific differences involve not only the number of simple sequence repeats, but also interruptions by imperfect repeats within a tandem-repeat array. Similar observations have previously been reported by other research workers (Garza et al., 1995; Estoup et al., 1995; FitzSimmons et al., 1995; Angers and Bernatchez, 1997). Also shown in Table 6.3, microsatellite alleles of identical size or pattern can be members of different genealogical clades as defined by the whole intron sequence as well as previous taxonomical classifications. For example, the A-subgenome intron of *G. barbadense* possesses (GAA)₆ which is identical to that of *G. nelsonii*, a G-genome species; whereas the A-subgenome in *G. hirsutum* possesses an imperfect microsatellite GGA(GAA)₈. It is also of interest to note that two accessions of *G. robinsonii* vary by a repeat unit, i.e. *robinsonii*-1 with 11 repeats, while *robinsonii*-2 with 10 repeats.

Table 6.3 A Summary of the Microsatellite (GAA)_n
Pattern in the pghD12-1 Intron from *Gossypium* Species

ACCESSION	GENOME	PATTERN
<i>hirsutum</i> (A)	AD	GGA(GAA) ₈
<i>barbadense</i> (A)	AD	(GAA) ₆
<i>herbaceum</i>	A	(GAA) ₁₀
<i>arboreum</i>	A	(GAA) ₈ GGA(GAA)
<i>barbadense</i> (D)	AD	(GAA) ₄
<i>raimondii</i>	D	(GAA) ₄
<i>gossypioides</i>	D	(GAA) ₂ GAGTAA(GAA) ₄
<i>trilobum</i>	D	(GAA) ₃ TAA(GAA) ₃
<i>australe-1</i>	G	(GAA) ₁₀ AAG(GAA) ₂
<i>australe-2</i>	G	(GAA) ₆
<i>australe-3</i>	G	(GAA) ₁₂
<i>nelsonii-1</i>	G	(GAA) ₆
<i>nelsonii-2</i>	G	(GAA) ₆
<i>bickii-1</i>	G	(GAA) ₅
<i>bickii-2</i>	G	(GAA) ₅
<i>bickii-3</i>	G	(GAA) ₅
<i>robinsonii-1</i>	C	(GAA) ₁₁
<i>robinsonii-2</i>	C	(GAA) ₁₀
<i>sturtianum-1</i>	C	(GAA) ₇
<i>sturtianum-2</i>	C	(GAA) ₇
<i>sturtianum-3</i>	C	(GAA) ₇
<i>anapoides</i>	K	(GAA) ₅ GCA(GAA) ₂
<i>costulatum</i>	K	(GAA) ₅ GCA(GAA) ₃
<i>cunninghamii</i>	K	(GAA) ₉
<i>enthyle</i>	K	(GAA) ₅ GCA(GAA) ₂
<i>exiguum</i>	K	(GAA) ₄ GAG(GAA) ₃
<i>londonderriense</i>	K	(GAA) ₁₀
<i>marchantii</i>	K	(GAA) ₉
<i>nobile</i>	K	(GAA) ₄ GCA(GAA) ₃
<i>pilosum</i>	K	(GAA) ₈
<i>populifolium</i>	K	(GAA) ₁₄
<i>pulchellum</i>	K	(GAA)TCA(GAA) ₇
<i>rotundifolium</i>	K	(GAA) ₄ GAG(GAA) ₃

Indeed, in contrast with their importance in intraspecific studies, microsatellites have yet to make any real contribution to phylogeny reconstruction (Kimura and Crow, 1964; Goldstein and Pollock, 1997). Although the reason for this is not yet entirely clear, part of the difficulty certainly stems from restrictions to divergence imposed by range constraints, irregularities and asymmetries in the mutation process, and the degradation of microsatellites over time (Goldstein and Pollock, 1997).

This result presented here on the molecular fine structure of microsatellite variation in the pghD12-1 intron among *Gossypium* species suggests the need for caution in the use of repeat counts at microsatellite loci as secure indicators of allelic relationships. On the other hand, despite the phylogenetic ambiguity at the interspecific level as shown in Table 6.3, the microsatellite (GAA)_n in the pghD12-1 intron may be useful for studying the intraspecific genetic relationships among populations within a *Gossypium* species. The rapid rate of evolution in microsatellites may provide reliable phylogenetic information for the different accessions in a particular species.

6.4.4 Alterations of DNA Sequence in the D-subgenome pghD12-1 Gene in Tetraploid Cottons

Since allotetraploid cotton is comprised of both A- and D-subgenomes, it is expected that there are two classes of sequences originating from the two different allotetraploid subgenomes, respectively, with each sequence sharing a high degree of sequence similarity to that of its corresponding putative subgenome donor. Moreover, in each allotetraploid species both types of introns should be detected in nearly equal proportions.

Initially, all the pghD12-1 introns amplified and cloned from *G. barbadense* and *G. hirsutum* were clearly more A-genome-like than D-genome-like. The same primers which successfully amplified the intron from D-genome diploids (*G. raimondii*, *G. gossypioides* and *G. trilobum*) failed to amplify the D-subgenome intron from each of the two allotetraploids. Eighteen independent PCR clones from each of the two

allotetraploids were partially sequenced in the region with distinct D-genome characteristics, but none of them was similar to D-genome intron sequence, e.g., that of *G. raimondii* or *G. trilobum* or *G. gossypoides*.

There are at least four possible explanations for this observation: (1) that the pghD12-1 gene has been deleted from the D-subgenome of allotetraploid cotton due to chromosome rearrangements; (2) that the intron of D-subgenome pghD12-1 gene has been lost; and (3) that interlocus concerted evolution has converted the DNA sequence of D-subgenome pghD12-1 gene to that of A-subgenome gene; (4) that the DNA sequence of D-subgenome pghD12-1 gene has undergone some dramatic sequence changes, especially in the primer regions and subsequently the primer mismatch in PCR resulted in the failure of PCR amplification.

The possibilities of the loss of the D-subgenome pghD12-1 gene can be ruled out by Southern blot analyses using pghD12-1-specific DNA probe. As described in Chapter Five, the DNA band representing the D-subgenome pghD12-1 gene could be identified in both allotetraploids, i.e. *G. barbadense* and *G. hirsutum*. Similar observation was made in the Southern blot analysis using an 1 kb EcoRI fragment of the pghD12-1 intron as a DNA probe showing the presence of introns of D-subgenome as well as the A-subgenome origins in allotetraploid cottons (data not shown). This also ruled out the possibility that the intron had been deleted from the D-subgenome pghD12-1 gene in allotetraploid.

Interestingly, when high-stringency conditions were applied in the Southern blot analysis with pghD12-1-specific DNA probe, the intensity of the band representing the D-subgenome pghD12-1 gene in allotetraploid cotton is obviously reduced in comparison with the A-subgenome band in the same lane (Fig. 6. 8). Since the D-genome diploid species *G. raimondii* remains strongly hybridised to the probe DNA and give constant strong signals, the apparent signal losses of D-subgenome bands in allotetraploid lanes clearly indicates that the pghD12-1 gene of the D-subgenome origin has undergone accelerated DNA sequence alterations, at least in the region homologous

to the probe DNA (3' UTR), following the formation of allotetraploids. It is apparent that the alteration of DNA sequence of D-subgenome intron was divergent away from the DNA sequence of A-subgenome pghD12-1 gene, rather than towards it. Therefore, the possibility of interlocus homogenisation converting the D-subgenome pghD12-1 gene to a A-subgenome gene can also be ruled out.

The sequences in UTR regions are normally regarded to be more amenable to change due to either the relaxed functional constraints or the requirement for functional divergence. Since the *d12PS1* primer was based on the DNA sequence in the 5' UTR, it is possible that the 5' UTR primer mismatch contributed to the failed amplification of the D-subgenome intron from allotetraploids. Indeed, PCR is a limited tool for investigation in that it can fail to detect some genes for which one primer has critical mismatches or in the case of competition between several copies. The PCR with a D-genome-specific primer (*d12DS1*) and primer *d12A4* shown in Fig. 6.6A confirmed that the D-subgenome intron does exist in the allotetraploid cottons and the mismatch of 5' UTR primer (*d12PS1*) was to blame for the unsuccessful amplification of D-subgenome intron. Further PCR experiments (Fig. 6.6B) demonstrated that the use of *d12it-1* primer in combination with *d12A4* primer was able to amplify the D-subgenome intron from *G. barbadense*, but not from *G. hirsutum*. Therefore, there are considerable sequence divergence in the D-subgenome introns of *G. barbadense* and *G. hirsutum* following their divergence. However, the partial DNA sequences of the D-subgenome introns from allotetraploid cottons showed no dramatic sequence alterations in comparison with the corresponding region in the introns of D-genome diploids (Fig. 6.7A). Therefore, until the isolation of full length D-subgenome pghD12-1 gene from allotetraploid cottons, the extent of the sequence alteration remains unknown.

The hybridisation of A- and D- genome progenitors and the subsequent polyploidisation that gave rise to allotetraploid cotton is a relatively recent event (1-2 MYA) in comparison with the divergence of A- and D- genome (6-11 MYA) and the divergence of Australian species from the rest of the *Gossypium* genus (20-30 MYA)

(Wendel and Albert, 1992). However, the successful amplification of the pghD12-1 intron from a wide range of *Gossypium* species using the same set of PCR primers and the relatively high sequence homology among these intron sequences is in sharp contrast to the D-subgenome intron in allotetraploids. This leads to the speculation that some evolutionary forces have acted on the D-subgenome pghD12-1 gene in allotetraploid cottons. Such a force may have caused either gene silencing or functional divergence of the D-subgenome pghD12-1 gene. It has been documented that the major gene duplication event, such as genome duplication *via* allopolyploidisation, can lead to gene redundancies which would resolve with the silencing of one member of a gene pair (see review by Pickett and Meeke-Wagner, 1995). The silenced or functionally diverged member of a gene pair, probably the D-subgenome pghD12-1 gene in this case, will have relatively rapid rates of mutation leading to the accumulation of dramatic sequence variations. Interestingly, the ORF portion of the available D-subgenome pghD12-1 sequence data (the first 48 amino acids) is well translated and highly conserved with the corresponding region in other *Gossypium* species. Therefore, it is possible that the D-subgenome pghD12-1 is still a functional gene.

6.4.5 Nuclear Introgression in Australian *Gossypium* Species

There are seventeen species native to Australia in the genus *Gossypium*. Collectively, these taxa comprise the C-, G- and K genomes (Fryxell 1965; 1979; 1992). C-genome comprises two species, *G. sturtianum* and *G. robinsonii*. *G. sturtianum* is widely distributed in central and southern Australia, whereas *G. robinsonii* has a relatively restricted range in western part of Western Australia. The G-genome contains three morphologically similar species: *G. australe*, *G. nelsonii* and *G. bickii*. All the rest twelve species belong to K-genome. The K-genome species have a distinctive geography, morphology and ecology, and exhibit features characteristic of 'fire-adapted' plants. Many of these species are poorly represented in collections and are perhaps the least understood part of the genus.

In contrast to the expectations based on this taxonomy, the maternally inherited chloroplast genome of G-genome species *G. bickii* was shown to be nearly identical to that of *G. sturtianum*, a morphologically distant C-genome species (Wendel et al., 1991). This discrepancy with traditional taxonomy was explained by a hypothesis that *G. bickii* has evolved with a biphyletic ancestry (Wendel and Albert, 1992). *G. sturtianum*, or a similar species, served as the maternal parent in an ancient hybridisation with a paternal donor from the lineage leading to *G. australe* and *G. nelsonii*.

However, morphological analysis has yielded no evidence to suggest that *G. bickii* is an intermediate between any C- and G-genome species. Instead, it shows a strong skewing towards its paternal progenitor lineage. All other source of evidence, including interspecific fertility relationships (Endrizzi et al., 1985; Stewart, 1995), parsimony analysis of nuclear markers with 70 variable allozymes and 13 restriction site mutations in ribosomal DNAs (Wendel et al., 1991), indicated that *G. bickii* shares a more recent common ancestor with its close morphological allies, *G. australe* and *G. nelsonii*, than it does with *G. sturtianum*. Since no *G. sturtianum* nuclear genes were detected in *G. bickii*, it was postulated that the nuclear genome of the maternal parent was eliminated from the hybrid or its descendent maternal lineage (Wendel et al., 1991; Wendel and Albert, 1992).

The present results from the phylogenetic analysis of pghD12-1 intron indicate all three accessions of *G. bickii* ranked next to the C-genome species, i.e. *G. robinsonii* and *G. sturtianum* rather than its G-genome sister taxa (Fig. 6.4). The sequence alignment in Fig. 6.3 also indicated that *G. bickii* intron is similar to one from a C-genome. According to Wendel and Albert (1992), *G. bickii* has biphyletic history and contains a G- nuclear genome and a C- chloroplast genome. Because the ω -6 desaturase gene is a nuclear gene, the *a priori* expectation is that *G. bickii* would appear as part of a monophyletic G-genome clade, as is the case for other nuclear markers (Wendel et al., 1991; Wendel and Albert, 1992). However, this is apparently not the case with the

current study of *pghD12-1* intron. From the present intron DNA sequence data, it is postulated that *G. bickii*'s capture of the C-genome chloroplast is also accompanied by nuclear introgression. To the best of my knowledge, this is probably the first evidence of nuclear introgression in *G. bickii*.

As shown in Fig. 6.4, an accession of *G. australe* (*australe-2*) appears to align more closely to *G. nelsonii*. There is a high degree of morphological resemblance between these two G-genome species. It is possible that some specimens of *G. nelsonii* were misidentified as *G. australe*, and *vice versa*. However, the careful re-examination of plant of this particular accession ruled out the possibility of sampling error. Thus, this observation perhaps represents a nuclear introgression between these two G-genome species. Since the other two accessions of *G. australe* differ from the *G. nelsonii* clade, it can be assumed that the nuclear introgression in *G. australe* observed here is an independent event.

Interspecific hybridisation and introgression are important evolutionary process in plants, although their full significance with respect to speciation at the diploid level remains unresolved (Nei, 1987). The classic situation for discovery of introgression is in hybrid zones between formerly allopatric, closely related taxa that are capable of interbreeding. The C-genome taxon *G. sturtianum* is morphologically variable and widely distributed across central Australia. It has been found sympatrically with species of all three genome groups. Natural hybridisation between *G. sturtianum* and *G. australe* has previously been reported, despite its rare occurrence (Fryxell, 1965).

As shown in Fig. 6.4, the K-genome sequences form a distinct group. The geographical distribution of the K-genome species are clearly limited, being restricted to the Kimberley region of northwestern Australia. The only exception is *G. cunninghamii* which occurs in the Cobourg Peninsula of the Northern Territory, approximately 500 km distant from other members of this genome. By analogue with the situation in *G. bickii*, the chloroplast genome of *G. cunninghamii* was also found to be donated by the *G. sturtianum* lineage (Wendel and Albert, 1992). However, the current study with

pghD12-1 intron ranked *G. cunninghamii* deep in the K-genome clade, as a sister lineage with *G. anapoides* and *G. londonderriense*. Hence the present study found no evidence of nuclear introgression in *G. cunninghamii*. Although the relationships among these K-genome species is outlined on the basis of pghD12-1 intron sequence, the K-genome species may share a more subtle evolutionary history which needs to be studied in the years to come.

6.4.6 The Phylogenetic Position of *G. gossypioides* pghD12-1 Intron

A D-genome diploid, *G. gossypioides*, is also worthy of a special mention here. *G. gossypioides* is narrowly distributed in Southern Mexico as a series of small scattered populations. In the current gene tree with pghD12-1 intron, *G. gossypioides* is embedded within the D-genome (Fig 6.4). This phylogenetic position is conventional in that it was expected from nearly all previous evidence, including comparative gross morphology (Fryxell, 1979, 1992), chromosome morphology and DNA content (Brown and Menzel, 1952a), hybrid fertility (Brown and Menzel, 1952b) and allozyme analysis (Wendel, referred by Wendel et al., 1995a).

In contrast to this unanimity of evidence, Wendel et al. (1995a) demonstrated that the ribosomal DNA sequences from *G. gossypioides* are unlike those of all other D-genome taxa, exhibiting extensive recombination with rDNA sequences from A-genome cottons. Subsequent to this finding, other repetitive DNAs from *G. gossypioides* have been found to be similarly introgressant (Zhao et al., 1996). Wendel et al. (1995a) attributed these unusual observations to an ancient hybridisation event, whereby *G. gossypioides* experienced contact with an A-genome, either at the diploid level, or at the triploid level as a result of hybridisation with a New World allopolyploid, followed by repeated backcrossing of the hybrid into the *G. gossypioides* lineage.

Although the extent and chromosomal distribution of introgressed sequence are not clear, it is likely that most of the single copy components of the D-nuclear genome have been restored by extensive backcrossing (Wendel et al., 1995a). Therefore, it is not

surprising that the pghD12-1 gene maintains its D-genome identity. However, a survey of terpenoids content in *Gossypium* species demonstrated that *G. gossypioides* synthesises four heliocides that are common in the two A-genome diploids and in all AD-genome allopolyploids, but are otherwise completely absent from D-genome diploids (Altman et al, 1990). Although the molecular genetics of the biosynthesis of these terpenoids is not well understood, the data may implicate the transfer of the relevant genes from A-genome to *G. gossypioides* during the hypothesised process of genome introgression. It is a intriguing question whether any selective significance exists owing to the functionality of a particular gene.

6.4.7 The Potential Gene Regulation by the pghD12-1 Intron

As it has been described in Chapter Five, the two cDNAs encoding microsomal ω -6 desaturase have distinct expression patterns. While the pghD12-1 is highly inducible during cottonseed embryo development, pghD12-2 has very low expression in both embryo and leaf tissues. Although the genomic clones corresponding to these two cDNAs have not been isolated, PCR results indicated that one of the major differences between these two genes is the presence of a large intron in the 5' UTR of the pghD12-1 gene.

Whether the gain or loss of such a large intron in cotton microsomal ω -6 desaturase genes can have significant effect on gene expression commands vast interest because of the potential impact on the efficient genetic manipulation of fatty acid desaturation by antisense or cosuppression technology.

A positive effect of introns on gene expression has been observed for many plant genes. Expression of reporter genes under the control of the maize *Adh1*, *Sh1*, *Bz1* or *Act* promoter is increased up to several hundred folds by the inclusion of an intron (Callis et al., 1987; Maas et al., 1991; Oard et al, 1989; Vasil et al., 1989). Similarly, *A. thaliana* genes encoding polyubiquitin (Norris et al., 1993), transcription factor EF-1 α (Currie

et al., 1991; 1993), and a protein kinase (Zhang et al., 1994) all have an intron in the 5' untranslated region that increases the expression of reporter gene fusions 2.5 to 1000 fold relative to intron-less controls. However, the mechanism of enhancement remains unclear, and may not be the same for all genes. Several plant introns have been indirectly shown to increase expression by a mechanism related to intron splicing *per se*. In these cases, the introns must be included within the transcribed portion of a gene and in the proper orientation in order to increase mRNA accumulation (Callis et al., 1987; Vasil et al., 1989). There is also a strong correlation between the ability of the first intron of the maize *Adh1* gene to be spliced and its capacity to enhance gene expression. Mutagenesis studies indicate that the sequences within an intron can be significantly modified by deletions or substitutions without interfering with the enhancement effect, as long as the intron can be efficiently spliced (Lüehrsen and Walbot, 1991, 1994). These observations suggest that the process of intron splicing is mechanistically related to the production of stable mRNA, possibly through coupling to some other aspects of mRNA maturation such as capping, polyadenylation, or export from the nucleus. However, there is also other reports that some enhancer elements were identified. For example, the first intron of *A. thaliana* elongation factor-1b contains an enhancer-like element and its enhancing effect was independent of its location with respect to the transcription start site (Gidekel et al., 1997).

The intron-GUS fusion construct described in this chapter (Fig. 6.9) is designed for a preliminary examination of the possible enhancing effect by this large intron present in the 5' UTR of *pghD12-1* gene. If the proposed enhancing effect is proved, further investigations towards the understanding of its working mechanism would be necessary. The determination of either the post-transcriptional factors or enhancer elements which influence its enhancing effects should provide strategies for high level expression of transgenes.

6.5 CONCLUSIONS

This chapter describes the isolation and characterisation of a large 5' UTR intron from pghD12-1 gene from 24 *Gossypium* species. For comparison, the corresponding introns from *A. thaliana* and *G. max* were also isolated and sequenced. Phylogenetic relationships among these *Gossypium* species on the basis of this intron are largely congruent with their traditional taxonomic, cytogenetic classifications and modern molecular phylogenetics revealed by analysis of chloroplast DNA and rDNA data. However, it also revealed some interesting observations in the aspect of evolutionary history of pghD12-1 gene in *Gossypium* species. Firstly, the observation of a selective rapid sequence alteration in the D-subgenome pghD12-1 gene in allotetraploids demonstrated the interesting phenomenon that evolutionary forces appear to be working on the redundant gene in allotetraploid cottons. Secondly, the first evidence of nuclear introgression in *G. bickii* by C-genome has been found. Thirdly, nuclear introgression by its sister lineage, *G. nelsonii*., is also detected in one accession of *G. australe*. Fourthly, this intron phylogenetic tree also sketched an outline for the poorly understood K-genome species. Although this type of intron sequences has never been utilised for phylogenetic purposes the results presented here suggest that it is a consistent and valuable complement to rDNA and its ITS sequences.

Chapter Seven

General Discussion

7.1 GENETIC ENGINEERING OF COTTONSEED OIL

Cottonseed produces highly unsaturated C18 oils, with 16-20% of oleate and 48-58% linoleate. Palmitate (20-26%) represents most of the saturated fatty acid composition in cottonseed oil (Chapter Three). During the *de novo* fatty acid biosynthesis in cottonseed, the majority of fatty acid chains extend up to stearyl-ACP and are then desaturated to oleoyl-ACP by a $\Delta 9$ stearyl-ACP desaturase. Very little stearyl-ACP is cleaved by the $\Delta 9$ stearyl-ACP thioesterase and therefore stearate exists only in small amounts in cottonseed oil. The high proportion of the only polyunsaturated fatty acid, linoleic acid, also indicates the strength of the microsomal ω -6 fatty acid desaturase in cotton. Therefore, $\Delta 9$ stearyl-ACP desaturase and microsomal ω -6 desaturase are the major fatty acid desaturase enzymes active in cottonseed. The cDNAs encoding these enzymes have been isolated from a cottonseed cDNA library and their high expression during storage lipid synthesis was confirmed by Northern blot analysis (Chapters Four and Five). These cDNAs are available for the genetic engineering of the fatty acid composition of cottonseed oil in order to improve quality and produce specialty oil types.

7.1.1 High Stearic Cottonseed Oil

Theoretically, high-stearic cottonseed oil can be produced by the suppression of the predominant gene *pgh9-1* coding for $\Delta 9$ stearoyl-ACP desaturase by sense co-suppression or antisense suppression. The decrease in $\Delta 9$ stearoyl-ACP desaturase levels will result in a decrease in the conversion of stearoyl-ACP to oleoyl-ACP and therefore allow the 18:0-ACP pool concentration to increase, facilitating increased flux through the stearoyl-ACP thioesterase. The gene constructs with both sense and antisense of full length cDNA, *pgh9-1*, driven by a soybean lectin promoter have been made and gene transformation by means of *Agrobacterium tumefaciens* cocultivation will be carried out using a routine procedure established in the CSIRO Plant Industry.

The potential for increasing stearic acid accumulation in cottonseed oil may also be achieved by the overexpression of the stearoyl-ACP thioesterase. An increase in stearate release from the ACP before stearoyl-ACP desaturase can act, may also effectively reduce the conversion of saturated to unsaturated fatty acid. The function of a $\Delta 9$ stearoyl-ACP thioesterase becomes even more important when the large amount of stearoyl-ACP is accumulated in plastids as a result of the suppression of $\Delta 9$ stearoyl-ACP desaturase. Therefore, it is postulated that the combination of $\Delta 9$ stearoyl-ACP desaturase and thioesterase would substantially increase the stearate content of cottonseed oil.

Another strategy to increase the stearate content is to enhance its synthesis by overexpression of KASII which catalyses the elongation of palmitic acid to stearic acid. The cDNAs encoding KASII enzyme have been isolated from several plant species, including barley (Wissenbach et al., 1992), *Cuphea* (Slabaugh et al., 1993) and jojoba (Lassner et al., 1996). Although the overexpression of a *B. napus* KASII cDNA in soybean had no effect on the final ratio of palmitic acid to stearic acid in soybean oil (Kinney, 1996), it is not clear whether KASII enzyme is present in functional excess in the fatty acid synthesis pathway. Furthermore, the concomitant overexpression of

several key enzymes in the FAS pathway might be required in order to achieve elevated stearate content in cottonseed oil.

7.1.2 Low Palmitic Cottonseed Oil

Although palmitate is one of the important constituents in cocoa butter, its consumption has also been associated with increased risk of coronary disease in humans (Khosla and Hayes, 1996). Thus, one of the desired traits in cottonseed oil is reduced palmitic acid content. Such a goal has not been previously achieved and it is not immediately clear what would be best way to tackle it. Nevertheless, strategies of choice may include: either decreasing the amount of palmitate leaving the plastid by suppressing the palmitoyl-ACP thioesterase; or increasing the conversion of palmitoyl-ACP to stearoyl-ACP by increasing the KASII activity. It is also possible that the palmitate content of oil is controlled during triacylglycerol assembly. As it will be discussed in the following section 7.2.1, a more recent study with site-directed mutagenesis of $\Delta 9$ stearoyl-ACP desaturase in castor bean has created a mutant enzyme, L118F/P179I, which has a strong substrate preference for palmitoyl-ACP (Cahoon et al., 1997b). This indicates a novel opportunity to reduce the palmitate content in vegetable oil in the future.

7.1.3 High Oleic Cottonseed Oil

Production of vegetable oil with high oleic acid content at the expense of polyunsaturated fatty acid has been a major goal in many lipid research laboratories around the world. Recent success of high oleic soybean oil was achieved using cosuppression of a microsomal ω -6 fatty acid desaturase (Kinney, 1996). Since the equivalent gene from cotton has been cloned as a result of the current study (Chapter Five), it is now possible to adopt the same strategy to develop a high oleic cottonseed oil. Among the two microsomal ω -6 fatty acid desaturase cDNAs obtained from cottonseed cDNA library, pghD12-1 is highly expressed concomitant to the lipid biosynthesis and therefore is the target gene for suppression.

7.1.4 Cocoa Butter Substitute

In future experiments with the suppression of *pgh9-1* and *pghD12-1* genes, it is reasonable to predict that the transformants obtained will be a population with different levels of stearic and oleic acid composition because of the positional effect of the transgenes. Besides the high stearic, high oleic genotypes, it is also desirable to achieve a genotype with its fatty acid composition in seed oil similar to that of cocoa butter. This can probably be obtained directly from the transformant populations through careful selection and evaluation, since the palmitate and oleate content in cottonseed oil is not too far away from the targeted percentage in cocoa butter. The downregulation of any of the two fatty acid desaturase genes alone would dramatically alter the components of the flux in fatty acid synthesis pathway and therefore affect the effectiveness of the other fatty acid desaturase. However, there is also a possibility that the cocoa butter-like genotype have to be achieved by combining high stearic and high oleic genotypes through crossing and selection. It may require a large transgenic population and breeding program to achieve such a genotype with equal composition of three major fatty acids, i.e. palmitic, stearic and oleic acids in seed oil.

7.1.5 A Single Dominant Fatty Acid in Cottonseed Oil

For the purpose of industrial applications, it is also ideal to obtain a genotype with very high percentage, e.g. 90% of a single fatty acid. However, it has been well established that the triacylglycerol positions are not filled randomly and that saturated fatty acids are usually excluded from the middle position in the triacylglycerols. This specificity is conferred by lysophosphatidic acid acyltransferase (LPAAT) (Browse and Somerville, 1991; Jones et al., 1995). Theoretically, the maximum level of a saturated fatty acid would not exceed 66%. However, there are plants in nature, such as castor bean, whose seed oil comprises up to 90% of a single fatty acid. This indicates that there are probably no basic physiological reasons preventing oilseeds from accumulating 80-90% of a particular fatty acid in their oil. The recent success with the high-lauric rapeseed

involved the introduction of a second exogenous gene encoding the LPAAT from coconut besides a lauroyl-ACP thioesterase from California Bay, which increased levels of lauric acid up to 60% in the first generation of transgenic plants (Voelker, 1995). Therefore, as the understanding of the fatty acid biosynthesis and triacylglycerol assembly advances, it might be possible to achieve such a cottonseed genotype with a single dominant fatty acid by manipulating the activities of the relevant enzymes.

7.2 MODIFICATION OF FATTY ACID DESATURASE BY PROTEIN ENGINEERING

7.2.1 Modification of Substrate and Positional Specificities of $\Delta 9$ Stearoyl-ACP Desaturase

As discussed in Chapter One, besides the most widely occurring $\Delta 9$ stearoyl-ACP desaturase, there are several acyl-ACP desaturases in nature which have diverged functions in the production of various types of monounsaturated fatty acids. Based on amino acid sequence comparisons of acyl-ACP desaturases and on the three-dimensional information from the recently reported crystal structure of the castor bean $\Delta 9$ stearoyl-ACP desaturase (Lindqvist et al., 1996), mutant enzymes were synthesised *in vivo* through site-directed mutagenesis and the ability to rationally modify acyl-ACP desaturase activities was demonstrated (Cahoon et al., 1997a). By analogy to the $\Delta 9$ stearoyl-ACP desaturase in castor bean, the key nucleotides of cottonseed pgh9-1 gene determining the substrate chain length and double bond position may be located and mutated in order to create novel monounsaturated fatty acids in cottonseed oil. For example, the reduction of the palmitate content in cottonseed oil might be achieved by expressing a novel $\Delta 9$ palmitoyl-ACP desaturase derived from cotton $\Delta 9$ stearoyl-ACP desaturase (pgh9-1) by altering its substrate chain length specificity. Furthermore, by altering the double bond positional specificity of $\Delta 9$ palmitoyl-ACP desaturase from $\Delta 9$ to $\Delta 4$ position of a palmitic acid chain, in combination with the overexpression of KASII, it might be possible to produce petroselinic acid which has potential use in the

production of low caloric margarine and as a precursor of adipic acid for the manufacture of nylon (Murphy, 1992; Ohlrogge, 1994).

7.2.2 Potential New Functions of Microsomal ω -6 Desaturase

In contrast to the soluble $\Delta 9$ stearoyl-ACP desaturase, the detailed tertiary protein structure of membrane-bound desaturases, including the microsomal ω -6 fatty acid desaturase remains unknown. Nevertheless, as more DNA sequence data of the corresponding cDNAs/genes from the various plant species continue to emerge, a comparison of the deduced primary structure of the proteins is expected to reveal some functional domains important for their specificity, for substrate binding and even for protein-protein interactions. Moreover, the first fatty acid hydroxylase gene in higher plant isolated from castor bean has been shown to be a modified membrane-bound fatty acid desaturase with significant sequence identity with microsomal ω -6 desaturase (van der Loo et al., 1995). More recently, the first fatty acid acetylenase and the first epoxygenase from higher plants have been isolated from *Crepis alpina* and *C. palaestina*, respectively. These genes share a high degree of homology to each other and to the castor bean hydroxylase gene and plant microsomal ω -6 desaturase genes (Lee et al., 1998). This indicated that probably only moderate structural changes in the microsomal ω -6 desaturase gene could alter its function, i.e. becoming either a desaturase, hydroxylase, epoxygenase or an acetylenase. Therefore, by domain swapping or site directed mutagenesis, microsomal ω -6 fatty acid desaturases, such as pghD12-1 from cotton, may be engineered to produce epoxy, hydroxy and conjugated unsaturated fatty acids with novel features. Many of these fatty acids could be valuable in the chemical industry for the production of paints, varnishes, plasticisers, lubricants and polymers. Ultimately, as our understanding of desaturase biochemistry and gene structures advances, it should be possible to engineer fatty acid desaturases that are able to insert double bonds at specified position in acyl chains of particular lengths will be produced .

7.3 OPTIMISATION OF TRANSGENE REGULATION

There are many other issues to be addressed in the transgene system in addition to selecting the most effective genes to suppress or overexpress, such as selecting the most effective seed-specific promoter in order to achieve optimal expression of the transgene or an endogenous gene.

7.3.1 Use of a Seed Storage Protein Promoter

Promoters of seed storage protein genes have been shown to be most useful for genetic engineering of lipid biosynthesis since, in many oilseeds, lipid and seed storage protein syntheses take place in the same cells and at the same developmental stage. The two most successful cases of the genetic manipulation of fatty acid desaturase genes have used promoters from seed storage protein genes, i.e. napin promoter driving an antisense $\Delta 9$ stearoyl-ACP desaturase in canola (Knutzon et al., 1992b) and a soybean β -conglycinin promoter driving the microsomal ω -6 desaturase FAD2-1 in soybean (Kinney, 1995). As described in section 2.10 in Chapter Two, the sense and antisense binary gene constructs of pgh9-1 or pghD12-1 driven by a soybean lectin seed-specific promoter have been made. However, the accuracy and strength of the temporal and spatial regulations by such a heterologous promoter in cotton remains to be studied in stably transformed cotton in future. In the meanwhile, it is also necessary to consider the potential utilisation of a promoter from a cotton seed-storage protein gene or other genes involved in seed storage lipid synthesis, such as a ACP gene or even a fatty acid desaturase gene as it will be discussed in next section.

7.3.2 Isolation and Potential Utilisation of a Fatty Acid Desaturase Gene Promoter

In contrast to the extensive studies on the promoters from the seed storage protein genes, to date only a few studies have focused on promoters of genes involved in fatty acid biosynthesis (Bearson et al., 1994; de Boer et al., 1996; de Silva et al., 1992;

Slocombe et al., 1994; Bao et al., 1997). One consistent trend found in these promoters is that the highest expression levels were observed in rapidly growing tissue and developing seeds. However, very little is known about the molecular mechanisms regulating the rates of transcription of the genes involved in the fatty acid biosynthesis.

The 5' flanking region of the cotton $\Delta 9$ stearoyl-ACP desaturase gene contains some putative seed-specific elements and a region homologous to the distal promoter region of the cotton 2S albumin seed-storage protein. In future experiments, each of the successive 5' deletions of the Gh9-1G promoter will be fused with a GUS reporter gene. As a first approach to identify *cis* regulatory elements involved in the control of cotton $\Delta 9$ stearoyl-ACP desaturase gene expression, the effect of deletion/GUS fusion constructs will be analysed using transgenic *Arabidopsis* or tobacco. Although it is currently possible to transform cotton, the production and maintenance of the large number of transgenic cotton necessary for promoter studies is much more laborious and time consuming relative to using a model system such as *Arabidopsis* or tobacco which are commonly used in promoter studies. However, heterologous systems may not always show an expression consistent with that of the homologous plants. It is also becoming increasingly clear that the promoter-reporter fusions can be prone to artifactual expression that does not accurately reflect the *in vivo* regulation of the gene of interest. Many other factors, including the UTRs and introns are also involved in the gene regulations which will be discussed in the sections 7.3.4 and 7.3.5. When the promoter sequence is introduced into cotton, it will be interesting to observe the potential effects of transcriptional silencing caused by the interactions between homologous promoters. Such a promoter-mediated silencing could also serve as an alternative approach for the suppression of cotton $\Delta 9$ stearoyl-ACP desaturase gene.

It is postulated that the pghD12-1 gene is mainly responsible for the production of linoleic acid which makes up more than half of the total fatty acids in cottonseed oil. It is, therefore, reasonable to expect that such a gene may contain some promoter elements conferring its strength. However, the promoter from microsomal ω -6

desaturase gene has been neither described nor characterised so far. In fact, very little information is available about the gene structures of the membrane-bound fatty acid desaturase genes as a gene family. Preliminary screening of a cotton genomic DNA library with a DNA fragment consisting of the 5' portion of pghD12-1 has resulted in some positive lambda plaques (not included in this thesis) and the purification and identification might lead to the isolation of the 5' flanking region of pghD12-1 gene. The primary characterisation of this region can be carried out using a promoter-GUS fusion system as described above.

7.3.3 Identification of Transcription Factor and Environmental Influences

While the initial studies of promoters focus on the identification of domains involved in the control of gene expression levels, it is also important to identify transcription factors that bind to these elements. Furthermore, additional controls on the expression of fatty acid desaturase genes may also be identified by search for mutants with altered fatty acid metabolism under various environmental conditions.

By analogy to studies carried out in other organisms, it is likely that the control of plant fatty acid desaturase genes involves a complex array of *cis* and *trans* acting factors. For example, the expression of animal fatty acid biosynthetic genes are controlled by hormones such as insulin (Moustaid et al., 1994), by dietary fatty acids (Roder et al., 1994), by glucose levels (Pribuus et al., 1995). Recent studies on the higher plant fatty acid desaturase gene has also indicated that their expression is regulated by ABA (Zou et al., 1995; Slocombe et al., 1994), light (Nishiuchi et al., 1995; Horiguchi et al., 1996), temperature (Gibson et al., 1994; Williams et al., 1997), senescence (Fukuchi-Mizutani et al., 1995), wounding (Hamada et al., 1996) or infection (Trizzini et al., 1993; Gadea et al., 1996; Kirsch et al., 1997a, b). In cotton, a $\Delta 9$ stearyl-ACP desaturase gene was found to be induced in the *Verticillium*-infected cotton roots (Dr Bruce Lyon, personal communication).

It has also been proposed that the enzymes of fatty acid synthesis are coordinately regulated, i.e. global transcriptional signals may control expression of many or all of the genes in the fatty acid biosynthesis pathway in plants (Ohlrogge and Jaworski, 1997). The identification of such a global transcriptional control may provide an effective means towards the genetic manipulation of fatty acid composition in cottonseed oil and other vegetable oils.

7.3.4 Importance of UTRs in Transgene Expression

The crucial role of the UTRs in many post-transcriptional processes, such as mRNA export, localisation, degradation and translation is now generally acknowledged (Ingelbrecht et al., 1989; Sonenberg, 1994; Koziel et al., 1996). The primary sequence context surrounding the initiation codon, the presence of binding sites for specific protein or antisense RNAs and the length of the 5' UTR sequence can have significant influence on gene expression in various ways (Sonenberg, 1994). The 3' UTR are responsible for the differential concentration of the corresponding proteins in different subcellular regions (Macdonald and Struhl, 1988; Pesole et al., 1997). Some sequence elements affecting mRNA degradation rate are also located in the 3' UTR region (Sonenberg, 1994).

In order to achieve maximum suppression of the endogenous desaturase genes in cottonseeds, it may be necessary to examine the regulatory effects of the 5' and 3' UTR regions of the transgenes, i.e. *pgh9-1* and *pghD12-1*. The initial experiment for the evaluation of the effects of UTRs can be carried out by using the whole cDNA sequence in comparison with the coding region of gene only.

7.3.5 Potential Enhancing Effect of the *pghD12-1* Intron

Introns can influence levels of gene expression by housing enhancer (Schultz et al., 1991) or silencer (Ottavio et al., 1990) elements that modulate mRNA levels through

interactions with the promoter and/or other regulatory elements. Within the first intron of the fatty acid synthase (FAS) gene in chicken, sequence elements that mediate both positive and negative effects on transcription have been identified (Oskouian et al., 1997). Similarly, a negative regulator of transcription was also discovered within the first intron of the human FAS gene (Hsu et al., 1996). However, besides the specific effects of such *cis*-acting elements it has also been reported that the mere presence of an intron, irrespective of its sequence, can elevate gene expression levels (Palmiter et al., 1991; Choi et al., 1991; Duncker et al., 1997).

The 1.1 kb 5' UTR intron isolated from pghD12-1 gene in cotton has been inserted into a 35S-GUS chimeric gene construct behind the 35S promoter (Chapter Six) and will be used to transform *Arabidopsis* in future experiments. Since pghD12-1 is highly expressed in cottonseed developing embryos, it is unlikely this intron has any negative effects on gene expression in developing seeds. The potential enhancing effects of this intron will be first evaluated in a heterologous system, such as stably transformed *Arabidopsis*. Manipulation of the intron orientation, sequence substitution, serial sequence deletion or site-directed mutagenesis may help to identify the underlying mechanism and the determinants which may contribute to gene enhancement. The possible interactions between promoter and intron can also be studied by using combinations of this intron and promoters with different specificities and strengths. The intron-mediated enhancement of gene expression might represent a powerful tool to increase the expression of transgenes and therefore contribute significantly to the genetic engineering of fatty acid desaturation in cottonseed oil.

7.4 INTRON GAIN OR INTRON LOSS FROM THE MICROSOMAL ω -6 DESATURASE GENES IN COTTON

Among the two microsomal ω -6 fatty acid desaturase genes isolated from cotton, pghD12-1 contains a large 5' UTR intron, but pghD12-2 does not (Chapters Five and Six). Although there is no definite evidence indicating that pghD12-2 is an intronless

gene, it does not contain an intron in the position where the pghD12-1 intron was found.

There are two different scenarios applicable to the evolution of cotton microsomal ω -6 desaturase gene: either the ancestral gene contained such an intron, but it was lost from the pghD12-2 gene following the gene divergence; or the ancestral gene did not contain an intron which was later gained by the pghD12-1 gene during its separated evolution. Indeed, many introns are often considered as relics of assembly of different functional domains in a far past (Gilbert and Glynias, 1993), whereas the others have been inserted secondarily through transposition (Purugganan and Wessler, 1993). Convincing cases for intron insertions have been made for the genes coding for actin, tubulin, and triosephosphate isomerase, among others (Dibb and Newman, 1989; Weber and Kabsch, 1994; Logsdon et al., 1995; Kwiatowski et al., 1995; Bhattacharya and Weber, 1997). Intron loss is also a rather common event in evolution (Geiss et al., 1994; Häger et al., 1996; Drouin and de Sa, 1997). While the debate about the intron gain and intron loss theories goes on, the current study does not provide the answer to the origin and fate of the intron in microsomal ω -6 desaturase gene in cotton. However, it is interesting that the intron-containing pghD12-1 gene is highly expressed during cottonseed embryogenesis, whereas the intronless pghD12-2 gene is not induced in the embryos. Whether the intron-loss of pghD12-2 or the intron-gain of pghD12-2 have played a role in their divergence of gene expression patterns remains an interesting open question.

7.5 EVOLUTION OF FATTY ACID DESATURASE GENES IN ALLOTETRAPLOID COTTON

Based on the results of PCR amplification of the pghD12-1 intron from allotetraploids, *G. hirsutum* and *G. barbadense* and Southern blot analysis, it appears that some accelerated sequence alterations have occurred in the 5' and 3' UTR region of the pghD12-1 gene of D-subgenome origin in these allotetraploids (Chapter Six). This

raises an interesting question about the fate of the D-subgenome pghD12-1 gene in the evolution of allotetraploid cotton.

It is a common phenomenon that the number of genes expressed in a polyploid is sufficiently reduced to return the polyploid to a level of expression similar or identical to that of the diploid ancestor (Soltis and Soltis, 1993). If the D-subgenome pghD12-1 gene is silenced, it is expected that many random nucleotide mutations can accumulate. Alternatively, targeted regions such as UTRs and/or some regulatory motifs can be the subject of rapid alterations owing to the mechanism of accelerated evolution and such a gene subsequently acquires a different function or expression pattern. The possibility that the diverging D-subgenome pghD12-1 gene might have developed a new function is of particular practical and theoretical interest to evolutionary studies. It has been previously demonstrated that the redundant gene copy in polyploids can acquire a new function by encoding a polypeptide with a slightly different catalytic capacity (Bailey and Thye, 1975), pattern of expression (Ferris and Whitt, 1979; Maniatis et al., 1980), or other kinetic properties (Gottlieb and Greve, 1981).

However, the extent of sequence alteration and its impact on the functionality of the D-subgenome gene based on the current primary data is still inconclusive (Chapter Six). Further study should focus on the isolation of a D-subgenome-specific DNA fragment which is large enough to be used as a DNA probe in Northern blot analysis to reveal the temporal and/or spatial expression pattern of the D-subgenome pghD12-1 gene. The UTR regions which have extensive sequence alterations should be particularly targeted. The initial experiment can be carried out using a PCR methodology, such as the adaptor-primer mediated PCR, using genomic DNA and D-subgenome-specific primer which has been developed in the current study.

7.6 POTENTIAL UTILISATION OF INTRON SEQUENCES IN THE PHYLOGENETIC ANALYSIS OF *GOSSYPIMUM* SPECIES

In comparison to exons, the nucleotide sequence of the 5' UTR intron of pghD12-1 gene is more variable and appears to have evolved at a relatively higher rate. The corresponding intron sequence among *Gossypium*, *Glycine max* and *A. thaliana* has diverged to such an extent that no sequence similarities among them can be identified. However, within the genus *Gossypium*, the sequence conservation of this intron is relatively high. There are also some insertions/deletions and enough sequence divergence to be used for evolution and phylogenetic studies. It also revealed some phylogenetic incongruence which may reflect underlying evolutionary histories of the species involved, such as nuclear introgression and accelerated evolution of redundant gene copies in allotetraploids. However, the current analysis was based on the "Corrected Distances" only and more detailed parsimonious phylogenetic analysis will be necessary in the future in order to obtain more robust resolution. Future studies may also include other genomes in *Gossypium* species, such as those from the African B- and F-genomes. The suitability of this intron at higher taxonomical level, such as the cotton tribe (*Malvaceae*) needs to be examined. As more sequence data accumulate, not only will the picture of the evolution of the microsomal ω -6 fatty acid desaturase gene in cotton be clearer, but also it will add to the growing list of nuclear gene trees which will reveal more evolutionary details in the cotton tribe as a whole.

REFERENCES

- Abián, J., Gelpi, E. and Pagé, M. (1991). Effect of abscisic acid on the linoleic acid metabolism in developing maize embryos. **Plant Physiol.** 95: 1277-1283.
- Akagi, H., Baba, T., Shimada, H. and Fujimura, T. (1995). Nucleotide sequence of a stearyl-acyl carrier protein desaturase cDNA from developing seeds of rice. **Plant Physiol.** 108: 845-846.
- Allen, E., Johnson, A.R., Fogerty, A.C., Pearson, J.A. and Shenstone, F.S. (1967). Inhibition by cyclopropane fatty acids of the desaturation of stearic acid in hen liver. **Lipids** 2: 419-423.
- Altman, D.W., Stipanovic, R.D. and Bell, A.A. (1990). Terpenoids in foliar pigment glands of A, D, and AD genome cottons: Introgression potential for pest resistance. **Theor. Appl. Genet.** 81: 447-454.
- Angers, B. and Bernatchez, L. (1997). Complex evolution of salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. **Mol. Biol. Evol.** 14: 230-238.
- Aronel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H.M. and Somerville, C.R. (1992). Map-based cloning of a gene controlling Omega-3 fatty acid desaturation in *Arabidopsis*. **Science** 258: 1353-1355.
- Ashton, A.R., Jenkins, C.L. and Whitfield, P.R. (1994). Molecular cloning of two different cDNAs for maize acetyl CoA carboxylase. **Plant Mol. Biol.** 24: 35-49.
- Auld, D.L., Heikkinen, M.K., Erickson, D.A., Sernyk, J.L. and Romero, J.E. (1992). Rapeseed mutants with reduced levels of polyunsaturated fatty acids and increased levels of oleic acid. **Crop Sci.** 32: 657-662.
- Baerson, S.R. and Lamppa, G.K. (1993). Developmental regulation of an acyl carrier protein gene promoter in vegetative and reproductive tissues. **Plant Mol. Biol.** 22: 255-267.
- Baerson, S.R., Heiden, V.M.G. and Lamppa, G.K. (1994). Identification of domains in an *Arabidopsis* acyl carrier protein gene promoter required for maximal organ-specific expression. **Plant Mol. Biol.** 26: 1947-1959.

- Bafor, M., Stobart, A.K. and Stymne, S. (1990). Properties of the glycerol acylating enzymes in microsomal preparations from the developing seeds of safflower (*Carthamus tinctorius*) and turnip (*Brassica campestris*) and their ability to assemble cocoa-butter type fats. **JAOCS** 67: 217-225.
- Bailey, A.E. (1948) **Cottonseed and Cottonseed Products**. Wiley, New York.
- Bailey, G. and Thye, L. (1975). Gene duplication in salmonid fishes: evolution of a lactate dehydrogenase with an altered function. In: Markert, C.L. (Ed.), **Isoenzymes, Genetics and Evolution** (pp. 401-414). Academic Press, New York.
- Baize, J.C. (1997). Oilseed industry's evolution and future. **INFORM** 8: 498-510.
- Baldoni, L., Georgi, L.L. and Abbott, A.G. (1996). Nucleotide sequence of a cDNA clone from *Olea europaea* encoding a stearyl-acyl carrier protein desaturase (Accession No. U58141). **Plant Physiol.** 111: 1353.
- Baldwin, B.G., Sanbderon, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S. and Donoghue, M.J. (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. **Ann. Missouri Bot. Gard.** 82: 247-277.
- Bao, X., Shorrosh, B.S. and Ohlrogge, J.B. (1997). Isolation and characterisation of an *Arabidopsis* biotin carboxylase gene and its promoter. **Plant Mol. Biol.** 35: 539-550.
- Bäumlein, H., Miséra, S., Luerben, H., Kölle, K., Horstman, C., Wobus, U. and Müller, A.J. (1994). The FUS3 gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis. **Plant J.** 6: 379-387.
- Bäumlein, H., Wobus, U., Pustell, J. and Kafatos, F. (1986). The legumin gene family: structure of a B-type gene of *Vicia faba* and a possible legumin gene specific regulatory element. **Nucl. Acids Res.** 14: 2707-2720.
- Bhattacharya, D. and Weber, K. (1997). The actin gene of the glaucocystophyte *Cyanophora paradoxa*: analysis of the coding region and introns, and an actin phylogeny of eukaryotes. **Curr. Genet.** 31: 439-446.
- Bhella, R.S. and MacKenzie, S.L. (1995). Nucleotide sequence of a cDNA from *Limnanthes douglasii* L. encoding a Δ -15 linoleic acid desaturase. **Plant Physiol.** 108: 861.

- Bothmann, H., Klein, B., Voetz, M., Pawlowski, K., Bautor, J., Schell, J. and Töpfer, R. (1993). Isolation and characterisation of a gene from *Cuphea lanceolata* encoding a β -ketoacyl-(ACP)-synthase I. **Biol. Chem. Hoppe-Seyler** 374: 521.
- Bourque, J.E. (1995). Antisense strategies for genetic manipulations in plants. **Plant Sci.** 105: 125-149.
- Bremer, K. (1988). The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. **Evolution** 42: 795-803.
- Brown, J.W.S., Smith, P. and Simpson, C.G. (1996). *Arabidopsis* consensus intron sequences. **Plant Mol. Biol.** 32: 531-535.
- Brown, L.C. and Kurtz, E.B. (1959). The *in vitro* synthesis of fats in cottonseed. **Agronomy J.** 51: 49-50.
- Brown, M.E. and Menzel, M.Y. (1952a). The cytology and crossing behaviour of *Gossypium gossypioides*. **Bull. Torrey Bot. club** 79: 110-125.
- Brown, M.E. and Menzel, M.Y. (1952b). Additional evidence on the crossing behaviour of *Gossypium gossypioides*. **Bull. Torrey Bot. Club** 79: 285-292.
- Browse, J. and Somerville, C. (1995). Glycerolipids. In: Meyerowitz, E. and Somerville, C. (Eds.), *Arabidopsis* (pp. 881-912). Cold Spring Harbor Laboratory Press, New York.
- Brubaker, C.L. and Wendel, J.F. (1993). On the specific status of *Gossypium lanceolatum* Todaro. **Genet. Resources Crop Evol.** 40: 165-170.
- Burden, W.R. (1995). Sources of lipids for the Australian food supply. **Supplement to Food Australia** 47: S18-S19.
- Bustos, M.M., Begum, D., Kalkan, F.A., Battraw, M.J. and Hall, T.C. (1991). Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. **EMBO J.** 10: 1469-1479.
- Bustos, M.M., Guitinan, M.J., Jordano, J., Begum, D., Kalkan, F.A. and Hall, T.C. (1989). Regulation of β -glucuronidase expression in transgenic tobacco plants by an A/T-rich, *cis*-acting sequence found upstream of a French bean β -phaseolin gene. **The Plant Cell** 1: 839-853.

- Caddle, M.S., Lussier, R.H. and Heintz, N.H. (1990). Intramolecular DNA triplexes, bent DNA and DNA unwinding elements in initiation region of an amplified dihydrofolate reductase repicon. **J. Mol. Biol.** 211: 19-33.
- Cahoon, E.B. and Ohlrogge, J.B. (1994). Metabolic evidence for the involvement of a $\Delta 4$ -palmitoyl-acyl carrier protein desaturase in petroselinic acid synthesis in coriander endosperm and transgenic tobacco cells. **Plant Physiol.** 104: 827-837.
- Cahoon, E.B., Becker, C.B., Shanklin, J. and Ohlrogge, J.B. (1994). cDNAs for isoforms of the $\Delta 9$ -stearoyl-acyl carrier protein desaturase from *Thunbergia alata* endosperm. **Plant Physiol.** 106: 807-808.
- Cahoon, E.B., Coughlan, S.J. and Shanklin, J. (1997a). Characterisation of a structurally and functionally diverged acyl-acyl carrier protein desaturase from milkweed seed. **Plant Mol. Biol.** 33: 1105-1110.
- Cahoon, E.B., Lindqvist, Y., Schneider, G. and Shanklin, J. (1997b). Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position. **Proc. Natl. Acad. Sci. USA** 94: 4872-4877.
- Cahoon, E.B., Shanklin, J. and Ohlrogge, J.B. (1992). Expression of a coriander desaturase results in petroselinic acid production in transgenic tobacco. **Proc. Natl. Acad. Sci. USA** 89: 11184-11188.
- Callis, J., Fromm, M. and Walbot, V. (1987). Introns increase gene expression in cultured maize cells. **Genes & Development** 1: 1183-1200.
- Carter, F.L., Castillo, A.E., Frampton, V.L. and Kerr, T. (1966). Chemical composition studies of seeds of the genus *Gossypium*. **Phytochemistry** 5: 1103-1112.
- Cavener, D.R. and Ray, S.C. (1991). Eukaryotic start and stop translation sites. **Nucl. Acids Res.** 19: 3185-3192.
- Chen, J. and Post-Beittenmiller, D. (1996). Molecular cloning of a cDNA encoding 3-ketoacyl-acyl carrier protein synthase III from leek. **Gene** 182: 45-52.
- Chen, L.J. and Moon, Y. (1995). Cloning and sequence of a cDNA encoding stearoyl-acyl-carrier-protein desaturase (accession No. L34346) from *Glycine max*. **Plant Physiol.** 109: 1498.
- Chen, Z.L., Pan, N.S. and Beachy, R.N. (1988). A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. **EMBO J.** 7: 297-302.

- Cherry, J.P. (1983). Cottonseed oil. **JAOCs** 60: 360-367.
- Cherry, J.P. and Berardi, L.C. (1983). Cottonseed. **CRC Handbook of Processing and Utilisation in Agriculture**. Boca Raton, Fla., CRC Press, Inc. 187-256.
- Cherry, J.P. and Leffler, H.R. (1984). Seed. In: Kohel, R.J. and Lewis, C.F. (Eds.), **Cotton** (pp. 511-569). American Society of Agronomy, Inc., Crop Science Society of America, Inc. and Soil Science Society of America, Inc. Publishers, Madison, Wisconsin, USA.
- Cherry, J.P., Katterman, F.R.H. and Endrizzi, J.E. (1970). Comparative studies of seed proteins of species of *Gossypium* by gel electrophoresis. **Evolution** 24: 431-447.
- Cherry, J.P., Kohel, R.J., Jones, L.A. and Powell, W.H. (1981). Cottonseed quality: Factors affecting feed and food uses. **Proceedings, Beltwide Cotton Production Conference**, Memphis, TN, National Cotton Council.
- Cherry, J.P., Simmons, J.G. and Kohel, R.J. (1978). Cottonseed composition of national variety test cultivars grown at different Texas locations. **Proc. Beltwide Cott. Prod. Res. Conf.** Dallas, Texas, Natl. Cotton Council. 47-50.
- Cherry, J.P., Simmons, J.G., Kohel, R.J., Cooper, H.B., Lehman, M., Dobbs, J., Fry, K.E., Kittock, D.L. and Henneberry, T.J. (1979). Relationship of cottonseed quality to genetic and agronomic practices. **Cott. Gin. Oil Mill Press** 80: 18-21.
- Cho, M.-J., Widholm, J.M. and Vodkin, L.O. (1995). Cassettes for seed-specific expression tested in transformed embryonic cultures of soybean. **Plant Molecular Biology Reporter** 13: 255-269.
- Choi, T., Huang, M., Gorman, C. and Jaenisch, R. (1991). A generic intron increases gene expression in transgenic mice. **Mol. Cell Biol.** 11: 3074.
- Cronn, R.C., Zhao, X., Paterson, A.H. and Wendel, J.F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. **J. Mol. Evol.** 42: 685-705.
- Currie, C., Axelos, M., Bardet, C., Atanassova, R., Chaubet, N. and Lescure, B. (1993). Molecular organisation and developmental activity of an *Arabidopsis thaliana* EF-1 α gene promoter. **Mol. Gen. Genet.** 218: 78-86.
- Currie, C., Liboz, T., Bardet, C., Gander, E., Médale, C., Axelos, M. and Lescure, B. (1991). *Cis*- and *trans*-acting elements involved in the activation of *Arabidopsis*

- thaliana* A1 gene encoding the translation elongation factor EF-1 α . **Nucl. Acids Res.** 19: 1305-1310.
- Das, N. and Ghosh, S. (1995). Molecular cloning and sequence analysis of a seed-expressed acyl carrier protein (ACP) gene from *Brassica campestris* (Agrani). **Biochimica et Biophysica Acta** 1258: 75-79.
- de Boer, G.J., Fawcett, T., Slabs, A., Nijkamp, J. and Stuije, A. (1996). Analysis of the *Arabidopsis* enoyl-ACP reductase promoter in transgenic tobacco. In: Williams, J.P., Khan, M.U. and Lem, N.W. (Eds.), **Physiology, Biochemistry and Molecular Biology of Plant Lipids**. Kluwer Academic Publishers, Dordrecht/Boston/London.
- de Pater, S.K., F., Kijne, J. and Chua, N-H. (1994). bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specifying element of the pea lectin promoter. **Plant J.** 6: 133-140.
- de Silva, J., Loader, N.M., Jarman, C., Windust, J.H., Hughes, S.G. and Safford, R. (1992). The isolation and functional characterisation of a *B. napus* acyl carrier protein 5' flanking region involved in the regulation of seed storage lipid synthesis. **Plant Mol. Biol.** 18: 1163-1172.
- de Silva, J., Loader, N.M., Jarman, C., Windust, J.H.C., Hughes, S.G. and Safford, R. (1990). The isolation and sequence analysis of two seed-expressed acyl-carrier protein genes from *Brassica napus*. **Plant Mol. Biol.** 14: 537-548.
- Dehesh, K., Jones, A., Knutzon, D.S. and Voelker, T.A. (1996). Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from *Cuphea hookeriana*. **The Plant Journal** 9: 167-172.
- De Joode, D.R. and Wendel, J.F. (1992). Genetic diversity and origin of the Hawaiian islands cotton, *Gossypium tomentosum*. **Am. J. Bot.** 79: 1311-1319.
- Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis progenitor programs for the VAX. **Nucl. Acids Res.** 12: 387-395.
- Dibb, N.J. and Newman, A.J. (1989). Evidence that introns arose at protosplice sites. **EMBO J.** 8: 2015-2021.
- Dickinson, C.D., Evans, R.P. and Nielsen, N.C. (1988). RY-repeats are conserved in the 5'-flanking regions of legume seed-protein genes. **Nucl. Acids Res.** 16: 371.

- Dinesh-Kumar, S.P. and Miller, W.A. (1993). Control of start codon choice on a plant viral RNA encoding overlapping genes. **The Plant Cell** 5: 679-692.
- Dörmann, P., Kridl, J.C. and Ohlrogge, J.B. (1994). Cloning and expression in *Escherichia coli* of a cDNA coding for the oleoyl-acyl carrier protein thioesterase from coriander (*Coriandrum sativum* L.). **Biochimica et Biophysica Acta** 1212: 134-136.
- Dörmann, P., Voelker, T.A. and Ohlrogge, J.B. (1995). Cloning and expression in *Escherichia coli* of a novel thioesterase from *Arabidopsis thaliana* specific for long-chain acyl-acyl carrier proteins. **Arch. Biochem. Biophys.** 316: 612-618.
- Doyle, J.J. (1992). Gene trees and species trees: molecular systematics as one-character taxonomy. **Syst. Bot.** 17: 144-163.
- Drouin, G. and de Sa, M.M. (1997). Loss of introns in the pollen-specific actin gene subfamily members of potato and tomato. **J. Mol. Evol.** 45: 509-513.
- Duncker, B.P., Davies, P.L. and Walker, V.K. (1997). Introns boost transgene expression in *Drosophila melanogaster*. **Mol. Gen. Genet.** 254: 291-296.
- Eccleston, V.S. and Harwood, J.L. (1990). Acylation reactions in developing avocado. In: Quinn, P.J. and Harwood, J.L (Eds.), **Plant Lipid Biochemistry, Structure and Utilisation** (pp. 178-180). Portland Press, London.
- Egli, M.A., Lutz, S.M. and Gengenbach, B.G. (1993). Cloning and sequence analysis of a maize acetyl-CoA carboxylase gene. **Plant Physiol.** (supplement) 102: 70.
- El-Nockrashy, A.S., Simmons, J.G. and Frampton, V.L. (1969). A chemical survey of seeds of the genus *Gossypium*. **Phytochemistry** 8: 1949-1958.
- El-Nokrashy, A.S., Mostafa, H.M., El-Fouly, M.M. and El-Shattory, Y. (1976). Biochemical changes in cottonseed during development and maturity. **Die Nahrung** 20: 125-132.
- Elborough, K.M., Swihoe, R., Winz, R., Kroon, J.T., Farnsworth, L., Fawcett, T., Martinez-Rivas, J.M. and Slabas, A.R. (1994). Isolation of cDNAs from *Brassica napus* encoding the biotin-binding and transcarboxylase domains of acetyl-CoA carboxylase: assignment of the domain structure in a full-length *Arabidopsis thaliana* genomic clone. **Biochem. J.** 301: 599-605.

- Elmore, C.D. and Leffler, H.R. (1976). Development of cotton fruit. III. Amino acid accumulation in protein and nonprotein nitrogen fractions of cottonseed. **Crop Sci.** 16: 867-871.
- Endrizzi, J.E., Turcotte, E.L. and Kohel, R.J. (1985). Genetics, cytology, and evolution of *Gossypium*. **Adv. Genet.** 23: 271-375.
- Espelund, M. and Jakobsen, K.S. (1992). Cloning and direct sequencing of plant promoters using primer-adaptor mediated PCR on DNA coupled to a magnetic solid phase. **BioTechniques** 13: 74-81.
- Estoup, A., Garnery, L., Solignac, M. and Cornuet, J.M. (1995). Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. **Genetics** 140: 679-695.
- Falcone, D.L., Gibson, S., Lemieux, B. and Somerville, C. (1994). Identification of a gene that complements an *Arabidopsis* mutant deficient in chloroplast $\omega 6$ desaturase activity. **Plant Physiol.** 106: 1453-1459.
- Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction fragments to high specific activity. **Anal. Biochem.** 132: 6-13.
- Ferandez-Martinez, J., del Rio, M. and de Haro, A. (1993). Survey of safflower (*Carthamus tinctorius* L.) germplasm for variants in fatty acid composition and other seed characters. **Euphytica** 69: 115-122.
- Ferris, S.D. and Whitt, G.S. (1979). Evolution of the differential regulation of duplicate genes after polyploidization. **J. Mol. Evol.** 12: 267-317.
- Filipowicz, W., Gniadkowski, M., Klahre, U. and Liu, H.X (1995). Pre-mRNA splicing in plants. In: Lamond, A. (Ed.), **Pre-mRNA Processing** (pp. 65-78). R.G. Landes Publishers, Georgetow, TX.
- Fincke, A. (1976). Untersuchungen zur fettsäurezusammensetzung von kakaobutter. **Gordian** 76: 324.
- Finkelstein, R.R., Tenbarge, K.M., Shumway, J.E. and Crouch, M.L. (1985). Role of ABA in maturation of rapeseed embryos. **Plant Physiol.** 78: 630-636.
- Fisher, G.S. and Cherry, J.P. (1983). Variation of cyclopropanoid fatty acids in cottonseed lipids. **Lipids** 18: 589-594.

- Fisher, G.S. and Schuller, W.H. (1983). Gas chromatographic analysis of cyclopropenoid acids in cottonseed oils. **JAACS** 58: 943-946.
- FitzSimmons, N.N., Moritz, C. and Moore, S.S. (1995). Conservation and dynamics of microsatellite loci over 300-million years of marine turtle evolution. **Mol. Biol. Evol.** 12: 432-440.
- Folch, J., Lees, M. and Stanley, G.H.S. (1957). A simple method for the isolation and purification of total lipids from animal tissues. **J. Biol. Chem.** 226: 497-.
- Fox, B.G., Shanklin, J., Somerville, C. and Munck, E. (1993). Stearoyl-acyl carrier protein $\Delta 9$ desaturase from *Ricinus communis* is a diiron-oxo protein. **Proc. Natl. Acad. Sci. USA** 90: 2486-2490.
- Fox, B.G., Shanklin, J., Ai, J., Loehr, T.M. and Sanders-Loehr, J. (1994). Resonance raman evidence for an Fe-O-Fe center in stearoyl-ACP desaturase. Primary sequence identity with other Diiron-oxo proteins. **Biochemistry** 33: 12776-12786.
- Fryxell, P.A. (1965). A revision of the Australian species of *Gossypium* with observation on the occurrence of *Thespesia* in Australia (*Malvaceae*). **Aust. J. Bot.** 13: 71-102.
- Fryxell, P.A. (1979). **The Natural History of the Cotton Tribe**. Texas A&M Univ. Press, College Station, TX.
- Fryxell, P.A. (1992). A revised taxonomic interpretation of *Gossypium* L. (*Malvaceae*). **Rheedea** 2: 108-165.
- Fukuchi-Mizutani, M., Savin, K., Cornish, E., Tanaka, Y., Ashikari, T., Kusumi, T. and Murata, N. (1995). Senescence-induced expression of a homologue of $\Delta 9$ desaturase in rose petals. **Plant Mol. Biol.** 29: 627-635.
- Gadea, J., Mayda, M.E., Conejero, V. and Vera, P. (1996). Characterisation of defense-related genes ectopically expressed in viroid-infected tomato plants. **Mol. Plant Microbe Interact.** 9: 409-415.
- Galau, G.A., Wang, H.Y-C. and Hughes, D.W. (1991). Sequence of the *Gossypium hirsutum* D-genome alloallele of Legumin A and its mRNA. **Plant Physiol.** 97: 1268-1270.

- Galau, G.A., Wang, H.Y.C. and Hughes, D.W. (1992a). Cotton *Lea4* (D19) and *LeaA2* (D132) group 1 *Lea* genes encoding water stress-related proteins containing a 20-amino acid motif. **Plant Physiol.** 99: 783-788.
- Galau, G.A., Wang, H.Y.C. and Hughes, D.W. (1992b). Cotton *Mat5-A* (C164) gene and *Mat5-D* cDNAs encoding methionine-rich 2S albumin storage proteins. **Plant Physiol.** 99: 779-782.
- Garza, J.C., Slatkin, M. and Freimer, N.B. (1995). Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. **Mol. Biol. Evol.** 12: 594-603.
- Gavel, Y. and von Heijne, G. (1990). A conserved cleavage-site motif in chloroplast transit peptides. **FEBS Lett.** 261: 455-458.
- Geiss, K.T., Abbas, G.M. and Makroff, C.A. (1994). Intron loss from the NADH dehydrogenase subunit 4 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate. **Mol. Gen. Genet.** 243: 97-105.
- Genez, A., McCarter, D., Nelson, J., Bleibaum, J.L., Thompson, G.A., Knauf, V.C. and Stalker, D. (1991). Molecular cloning and characterisation of β -ketoacyl-ACP synthase genes from *Ricinus communis*. In: **Proceedings of The 3rd International Congress**, The International Society for Plant Molecular Biology, Tuscon, Arizona, Oct. 6-11, 1991 Poster 719.
- Gibson, S., Arondel, V., Iba, K. and Somerville, C. (1994). Cloning of a temperature-regulated gene encoding a chloroplast ω -3 desaturase from *Arabidopsis thaliana*. **Plant Physiol.** 106: 1615-1621.
- Gidekel, M., Jimenez, B. and Herrera-Estrella, L. (1997). The first intron of the *Arabidopsis thaliana* gene coding for elongation factor 1 β contains an enhancer-like element. **Gene** 170: 201-206.
- Gilbert, W. and Glynias, M. (1993). On the ancient nature of introns. **Gene** 135: 137-144.
- Gipson, J.R. and Joham, H.E. (1969). Influence of night temperature on growth and development of cotton (*Gossypium hirsutum* L.) IV. Seed quality. **Agronomy J.** 61: 365-367.

- Giraudat, J., Parcy, F., Bertauche, N., Gost, F., Leung, J., Morris, P.C., Bouvier-Durand, M. and Vartanian, N. (1994). Current advances in abscisic acid action and signalling. **Plant Mol. Biol.** 26: 1557-1577.
- Goldberg, R.B., Barker, S.J. and Perez-Grau, L. (1989). Regulation of gene expression during plant embryogenesis. **Cell** 56: 149-160.
- Goldberg, R.B., de Paiva, G. and Yadegari, R. (1994). Plant embryogenesis: zygote to seed. **Science** 266: 605-614.
- Goldstein, D.B. and Pollock, D.D. (1997). Launching microsatellite: a review of mutation processes and methods of phylogenetic inference. **J. Heredity** 88: 335-342.
- Goodall, G.J. and Filipowicz, W. (1990). The minimum functional length of pre-mRNA introns in monocot and dicot plants. **Plant Mol. Biol.** 14: 727-733.
- Gornicki, P., Podkowinski, J., Scappino, L.A., DiMaio, J., Ward, E. and Haselkorn, R. (1994). Wheat acetyl-coenzyme A carboxylase: cDNA and protein structure. **Proc. Natl. Acad. Sci. USA** 91: 6860-6864.
- Gottlieb, L.D. and Greve, L.C. (1981). Biochemical properties of duplicated isozymes of phosphoglucose isomerase in the plant *Clarkia xantiana*. **Biochem. Genet.** 19: 155-172.
- Graef, G.L., Fehr, W.R. and Hammond, E.G. (1985). Inheritance of three stearic acid mutants of soybean. **Crop Sci.** 25: 1076-1079.
- Grayburn, W.S., Collins, G.B. and Hildebrand, D.F. (1992). Fatty acid alteration by a $\Delta 9$ desaturase in transgenic tobacco tissue. **Bio/Technology** 10: 675-678.
- Green, A.G. (1986a). A mutant phenotype of flax (*Linus usitatissimum* L.) containing very low levels of linolenic acid in its seed oil. **Can. J. Plant Sci.** 66: 499-503.
- Green, A.G. (1986b). Genetic control of polyunsaturated fatty acid biosynthesis in Flax (*Linum usitatissimum*) seed oil. **Theor. Appl. Genet.** 72: 654-661.
- Green, A.G. and Marshall, D.R. (1984). Isolation of induced mutants in linseed (*Linum usitatissimum*) having reduced linolenic acid content. **Euphytica** 33: 321-328.
- Grellet, F., Cooke, R., Raynal, M., Laudie, M. and Delseny, M. (1993). *Arabidopsis thaliana* systematic cDNA sequencing reveals a gene with homology with *Umbellularia californica*. **Plant Physiol. Biochem.** 31: 599-602.

- Griffiths, G., Brechany, E.Y., Christie, W.E., Stymne, S. and Stobart, A.K. (1989). Synthesis of octadecatetraenoic acid (OTA) in borage (*Borago officinalis*). In: Biacs, P.A., Gruiz, K and Kremmer, T. (Eds.), **Biological Role of Plant Lipids** (pp. 151-154). Plenum, New York.
- Gubler, U. and Hoffman, B.J. (1983). A simple and very efficient method for generating cDNA libraries. **Gene** 25: 263-269.
- Guiltinan, M.J., Marcotte, W.R. and Quantrano, R.S. (1990). A plant leucine zipper protein that recognises an abscisic acid response element. **Science** 250: 267-271.
- Gunstone, F.D. and Norris, F.A. (1983). **Lipids in Foods**. Pergamon Press, Oxford.
- Gunstone, F.D., Harwood, J.L. and Padley, F.B. (1986). **The Lipid Handbook**. Chapman and Hall, London.
- Gurr, M.I., Robinson, M.P. and James, A.T. (1969). Mechanism of formation of polyunsaturated fatty acids by photosynthetic tissue. **Eur. J. Biochem.** 9: 70-78.
- Häger, K.-P., Müller, B., Wind, C., Erbach, S. and Fischer, H. (1996). Evolution of legumin genes: loss of an ancestral intron at the beginning of angiosperm diversification. **FEBS Lett.** 387: 94-98.
- Hamada, T., Kodama, H., Nishimura, M. and Iba, K. (1994). Cloning of cDNA encoding tobacco omega-3 fatty acid desaturase. **Gene** 147: 293-294.
- Hamada, T., Nishiuchi, T., Kodama, H., Nishimura, M. and Iba, K. (1996). cDNA cloning of a wounding-inducible encoding a plastid ω -3 fatty acid desaturase from tobacco. **Plant Cell Physiol.** 37: 606-611.
- Hansen, L. (1987). Three cDNA clones for barley leaf acyl carrier proteins I and III. **Carlsberg Res. Commun.** 52: 381-392.
- Hansen, L. and Kauppinen, S. (1991). Barley acyl carrier protein II: nucleotide sequence of cDNA clones and chromosomal location of the Acl2 gene. **Plant Physiol.** 97: 472-474.
- Hansen, L. and von Wettstein-Knowles, P. (1991). The barley gene Acl1 and Acl3 encoding acyl carrier proteins I and III are located on different chromosomes. **Mol. Gen. Genet.** 229: 467-478.
- Harris, R.V. and James, A.T. (1965). Linoleic and alpha-linolenic acid biosynthesis in plant leaves and green alga. **Biochimica et Biophysica Acta** 106: 456-464.

- Harvey, B.L. and Downey, R.K. (1964). The inheritance of erucic acid content in rapeseed (*Brassica napus*). **Can. J. Plant Sci.** 44: 104-111.
- Harwood, J.L. (1988). Fatty acid metabolism. **Ann. Rev. Plant Physiol. Plant Mol. Biol.** 39: 101-138.
- Harwood, J.L. (1996). Recent advances in the biosynthesis of plant fatty acids. **Biochimica et Biophysica Acta** 1301: 7-56.
- Hedley, D. (1997). The future of the oils and fats market -an update on key trends for the 21st century. **Oils & Fats Markets** 13: 46-50.
- Heinkoff, S. and Dressen, T.D. (1989). *Trans*-inactivation of the *Drosophila* Brown Gene: evidence for transcriptional repression and somatic pairing dependence. **Proc. Natl. Acad. Sci. USA** 86: 6704-6708.
- Heinz, E. (1993). Biosynthesis of polyunsaturated fatty acids. In: Moore Jr., T.S. (Ed.), **Lipid Metabolism in Plants** (pp. 33-89). CRC Press
- Heppard, E.P., Kinney, A.J., Stecca, K.L. and Miao, G.H. (1996). Developmental and growth temperature relation of two different microsomal ω -6 desaturase genes in soybeans. **Plant Physiol.** 110: 311-319.
- Hitz, W.D., Yadav, N.S., Reiter, R.S., Mauvais, C.J. and Kinney, A.J. (1995). Reducing polyunsaturation in oils of transgenic canola and soybean. In: Kader, J.C. and Mazliak, P. (Eds.), **Plant Lipid Metabolism** (pp. 506-508). Kluwer Academic, Netherlands.
- Hlousek-Radojcic, A., Post-Beittenmiller, D. and Ohlrogge, J.B. (1992). Expression of constitutive and tissue-specific acyl carrier protein isoforms in *Arabidopsis*. **Plant Physiol.** 98: 206-214.
- Holbrook, L.A., Magus, J.R. and Taylor, D.C. (1992). Absciscic acid induction of elongase activity, biosynthesis and accumulation of very long chain monounsaturated fatty acids and oil body proteins in microspore-derived embryos of *Brassica napus* L. cv Peston. **Plant Sci.** 84: 99-115.
- Horiguchi, G., Iawkawa, H., Kodama, H., Kawakami, N., Nishimura, M. and Iba, K. (1996). Expression of a gene for plastid ω -3 fatty acid desaturase and changes in lipid and fatty acid compositions in light- and dark-grown wheat leaves. **Physiol. Plant.** 96: 275-283.

- Hsu, M.H., Chirala, S.S. and Wakil, S.J. (1996). Human fatty-acid synthase gene. **J. Biol. Chem.** 271: 13584-13592.
- Hughes, D.W., Wang, H.Y.C. and Galau, G.A. (1993). Cotton (*Gossypium hirsutum*) MatP6 and MatP7 Oleosin genes. **Plant Physiol.** 101: 697-698.
- Iba, K., Gibson, S., Nishiuchi, T., Fuse, T., Nishimura, M., Arondel, V., Hugly, S. and Somerville, C. (1993). A gene encoding a chloroplast omega-3 fatty acid desaturase complements alterations in fatty acid desaturation and chloroplast copy number of the fad7 mutant of *Arabidopsis thaliana*. **J. Biol. Chem.** 268: 24099-24105.
- Ichihara, K. (1984). Sn-glycerol-3-phosphate acyltransferase in a particulate fraction from maturing safflower seeds. **Arch. Biochem. Biophys.** 232: 685-689.
- Ichihara, K., Asahi, T. and Fuji, S. (1987). 1-Acyl-sn-glycerol-3-phosphate acyltransferase in maturing safflower seeds and its contribution to the non-random fatty acid distribution of triacylglycerol. **Eur. J. Biochem.** 167: 339-347.
- Ingelbrecht, L.W., Herman, L.M.F., Dekeyser, R.A., Montagu, M.C. and Depicker, A.G. (1989). Different 3' end regions strongly influence the level of gene expression in plant cells. **The Plant Cell** 1: 671-680.
- Ivanov, P., Petkov, D., D., Nikolova, V. and Petchev, E. (1988). Sunflower breeding for high palmitic acid content in the oil. **Proc. 12th Int. sunflower Conf.**, Nov Sad, Yugoslavia, Int. Sunflower Assoc. Toowoomba, Australia.
- Izawa, T., Foster, R. and Chua, N-H. (1993). Plant bZIP protein DNA binding specificity. **J. Mol. Biol.** 230: 1131-1144.
- Jackson, M.R., Nilsson, T. and Peterson, P.A. (1990). Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. **EMBO J.** 9: 3153-3162.
- James, D.W.J. and Dooner, H.K. (1990). Isolation of EMS-induced mutants in *Arabidopsis* altered in seed fatty acid composition. **Theor. Appl. Genet.** 80: 241-245.
- James, D.W.J., Lim, E., Keller, J., Plooy, I., Raiston, E. and Dooner, H.K. (1995). Directed tagging of the *Arabidopsis* FATTY ACID ELONGATION1 (FAE1) gene with the maize transposon activator. **The Plant Cell** 7: 309-319.

- Jaworski, J.G. (1987). Biosynthesis of monoenoic and polyenoic fatty acids. In: Stumpf, P.K. (Ed), **Lipids: Structure and Function** (pp. 159-174). Academic Press, Inc.
- Jaworski, J.G. and Stumpf, P.K. (1974). Properties of a soluble stearyl-acyl carrier protein desaturase from maturing safflower. **Arch. Biochem. Biophys.** 162: 158-165.
- Jaworski, J.G., Clough, R.C. and Barnum, S.R. (1989). A cerulenin insensitive short chain 3-ketoacyl-acyl carrier protein synthase in *Spinacia oleracea* leaves. **Plant Physiol.** 90: 41.
- Jennings, M.L. (1989). Topography of membrane proteins. **Annu. Rev. Biochem.** 58: 999-1027.
- Johnson, A.R., Pearson, J.A., Shenstone, F.S. and Fogarty, A.C. (1967). Inhibition of the desaturation of stearic to oleic acid by cyclopropene fatty acids. **Nature** 214: 1244-1245.
- Jones, A., Davies, H.M. and Voelker, T.A. (1995). Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. **The Plant Cell** 7: 359-371.
- Kajimoto, G., Yoshida, H., Shibahara, A. and Yamashoh, S. (1979). Changes in the compositions of lipids and fatty acids in cottonseeds during maturation. **Nippon Nogeikagaku Kaishi** 53: 317-320.
- Kashi, Y., Tikochinski, Y., Genislaw, E., Iragi, F., Nave, A., Beckmann, J.S., Gruenbaum, Y. and Soller, M. (1990). Large restriction fragments containing poly-TG are highly polymorphic in a variety of vertebrates. **Nucl. Acids Res.** 18: 1129-1132.
- Kater, M.M., Koningstein, G.M., Nijkamp, H.J. and Stuitje, A.R. (1991). cDNA cloning and expression of *Brassica napus* enoyl-acyl carrier protein reductase in *Escherichia coli*. **Plant Mol. Biol.** 17: 895-909.
- Kato, T., Shirano, Y., Kawazy, T., Tada, Y., Itoh and Shibata, D. (1991). A modified β -glucuronidase gene: sensitive detection of plant promoter activities in suspension-cultured cells of tobacco and rice. **Plant Mol. Biol. Rep.** 9: 333-339.

- Kearn, E.V., Hugly, S. and Somerville, C.R. (1991). The role of cytochrome b5 in Delta-12 desaturation of oleic acid by microsomes of safflower (*Carthamus tinctorius*). **Arch. Biochem. Biophys.** 284: 431-436.
- Khandjian, E.W. (1987). Optimised hybridisation of DNA blotting and fixed to nitrocellulose and nylon membranes. **Bio/Technology** 5: 165-167.
- Khattab, A.H., Khalifa, H. and El Tinay, A.H. (1977). Chemical composition of seeds of some species of the genus *Gossypium*. **J. Agri. Sci. Camb.** 88: 55-59.
- Khosla, P. and Hayes, K.C. (1996). Dietary *trans*-monounsaturated fatty acids negatively impact plasma lipids in humans: critical review of the evidence. **J. Am. Coll. Nutr.** 15: 325-339.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. **J. Mo. Evol.** 16: 111-120.
- Kimura, M. and Crow, J.F. (1964). The number of alleles that can be maintained in a finite population. **Genetics** 49: 725-738.
- King, E.E. and Leffler, H.R. (1979). Nature and patterns of proteins during cotton seed development. **Plant Physiol.** 63: 260-263.
- Kinney, A.J. (1995). Improving soybean seed quality. In: **Induced Mutations and Molecular Techniques for Crop Improvement** (pp. 101-113). IAEA and FAO, Vienna.
- Kinney, A.J. (1996). Designer oils for better nutrition. **Nature Biotechnology** 14: 946.
- Kinney, A.J. and Hitz, W.D. (1995). Improved soybean oils by genetic engineering. In: **The Proceedings of Biochem. Mol. Biol. Plant Fatty Acids Glycerolipids Symp.**, South Lake Tahoe, Calif.
- Kinney, A.J., Hirtz, W.D. and Yadav, N.S. (1990). Stearoyl-ACP desaturase and a β -ketoacyl-ACP acyl carrier protein synthase I isozyme from barley. In: Quinn, P.J. and Harwood, J.L. (Eds.), **Plant Lipid Biochemistry, Structure and Utilisation** (pp. 126-128). Portland Press, London.
- Kirsch, C., Hahlbrock, K. and Somssich, I.E. (1997a). Rapid and transient induction of a parsley microsomal Δ 12 fatty acid desaturase mRNA by fungal elicitor. **Plant Physiol.** 115: 283-289.

- Kirsch, C., Takamiya-Wik, M., Reinold, S., Hahlbrock, K. and Somssich, I.E. (1997b). Rapid, transient, and highly localised induction of plastidial ω -3 fatty acid desaturase mRNA at fungal infection sites in *Petroselinum crispum*. **Proc. Natl. Acad. Sci. USA** 94: 2079-2084.
- Klein, B., Pawlowski, K., Horicke-Grandpierre, C., Schell, J. and Töpfer, R. (1992). Isolation and characterisation of a cDNA from *Cuphea lanceolata* encoding a β -ketoacyl-ACP reductase. **Mol. Gen. Genet.** 233: 122-128.
- Knowles, P. (1989). Genetics and breeding of oil crops. In: Röbbelen, G., Downey, R.K. and Ashri, A. (Eds.), **Oil Crops of the World** (pp. 260-282). McGraw Hill, New York.
- Knutzon, D.S., Bleibaum, J.L., Nelsen, J., Kridl, J.C. and Thompson, G.A. (1992a). Isolation and characterisation of two safflower oleoyl-acyl carrier protein thioesterase cDNA clones. **Plant Physiol.** 100: 1751-1758.
- Knutzon, D.S., Lardizabal, K.D., Nelsen, J.S., Bleibau., J.L., Davies, H.M. and Metz, J.G. (1995). Cloning of a coconut endosperm cDNA encoding a 1-acyl-sn-glycerol-3-phosphate acyltransferase that accepts medium-chain-length substrates. **Plant Physiol.** 109: 999-1006.
- Knutzon, D.S., Scherer, D.E. and Schreckengost, W.E. (1991). Nucleotide sequence of a complementary DNA clone encoding stearoyl-acyl carrier protein desaturase from castor bean, *Ricinus communis*. **Plant Physiol.** 96: 344-345.
- Knutzon, D.S., Thompson, G.A., Radke, S.E., Johnson, W.B., Knauf, V.C. and Kridl, J.C. (1992b). Modification of *Brassica* seed oil by antisense expression of a stearoyl-acyl carrier protein desaturase gene. **Proc. Natl. Acad. Sci. USA** 89: 2624-2628.
- Kodama, H., Akagi, H., Kusumi, K., Fujimura, T. and Iba, K. (1997). Structure, chromosomal location and expression of a rice gene encoding the microsome ω -3 fatty acid desaturase. **Plant Mol. Biol.** 33: 493-502.
- Kohel, R.J. and Cherry, J.P. (1983). Variation of cottonseed quality with stratified harvest. **Crop Sci.** 23: 1119-1123.
- Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. **Cell** 44: 283-292.

- Kozak, M. (1989). Context effects and inefficient initiation at non-AUG codons in eukaryotic cell free translation systems. **Mol. Cell Biol.** 9: 5073-5080.
- Koziel, M.G., Carozzi, N.B. and Desai, N. (1996). Optimising expression of transgenes with an emphasis on post-transcriptional events. **Plant Mol. Biol.** 32: 393-405.
- Kwiatowski, J., Krawczyk, M., Kornacki, Bailey, K. and Ayala, F.J. (1995). Evidence against the exon theory of genes derived from the triosephosphate isomerase gene. **Proc. Natl. Acad. Sci. USA** 92: 8503-8506.
- Lamppa, G. and Jacks, C. (1991). Analysis of two linked genes coding for the acyl carrier protein (ACP) from *Arabidopsis thaliana* (Columbia). **Plant Mol. Biol.** 16: 469-474.
- Lassner, M.W., Lardizabal, K. and Metz, J.G. (1996). A jojoba β -ketoacyl-CoA synthase cDNA complements the canola fatty acid elongation mutation in transgenic plants. **The Plant Cell** 8: 281-292.
- Lassner, M.W., Levering, C.K., Davies, H.M. and Knutzon, D.S. (1995). Lysophosphatidic acid acyltransferase from meadowfoam mediates insertion of erucic acid at the sn-2 position of triacylglycerol in transgenic rapeseed oil. **Plant Physiol.** 109: 1389-1394.
- Lawhon, J.T., Cater, C.M. and Mattil, K.F. (1976). Evaluation of the food use potential of sixteen varieties of cottonseed. **JAOCs** 54: 75-80.
- Lee, M., Lenman, M., Banas, A., Bafar, M., Singh, S., Schweizer, M., Nilsson, R., Liljenberg, C., Dahlqvist, A., Gummesson, P-O., Sjö Dahl, S., Green, A. and Stymne, S. Identification of non-heme diiron proteins that catalyze triple bond and epoxy group formation. **Science** 280: 915-918.
- Lelievre, J.M., Oliveira, L.O. and Nielsen, N.C. (1992). 5'-CATGCAT-3' elements modulate the expression of glycinin genes. **Plant Physiol.** 98: 387-391.
- Lemieux, B., Miquel, M., Somerville, C.R. and Browse, J. (1990). Mutants of *Arabidopsis* with alterations in seed lipid fatty acid composition. **Theor. Appl. Genet.** 80: 234-240.
- Levinson, G. and Gutman, G.A. (1987). High frequency of short frameshifts in poly-CA/GT tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. **Nucl. Acids Res.** 15: 5323-5338.

- Lindqvist, Y., Huang, W., Schneider, G. and Shanklin, J. (1996). Crystal structure of $\Delta 9$ stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. **EMBO J.** 15: 4081-4092.
- Liu, H.X. and Filipowicz, W. (1996). Mapping of branchpoint nucleotides in mutant pre-mRNAs expressed in plant cells. **The Plant Journal** 9: 381-389.
- Loader, N.M., Woolner, E.M., Hellyer, A., Slabas, A.R. and Safford, R. (1993). Isolation and characterisation of two *Brassica napus* embryo acyl-ACP thioesterase cDNA clones. **Plant Mol. Biol.** 23: 769-778.
- Logsdon, J.M., Tyshenkio, M.G., Dixon, C., Jafari, J.D., Walker, V.K. and Palmer, J.D. (1995). Seven newly discovered intron positions in the triosephosphate isomerase gene: evidence for the intron-late theory. **Proc. Natl. Acad. Sci. USA** 92: 8507-8511.
- Lou, H., McCullough, A.J. and Schuler, M.A. (1993). 3' splice site selection in dicot plant nuclei is position independent. **Mol. Cell Biol.** 13: 4485-4493.
- Lu, Q., Wallrath, L.L., Granok, K. and Elgin, S.C.R. (1993). (CT)_n(GA)_n repeats and heat shock elements have distinct roles in chromatin structure and transcription activation of the *Drosophila* hsp26 gene. **Mol. Cell Biol.** 13: 2802-2841.
- Lüehrsen, K.R. and Walbot, V. (1991). Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. **Mol. Gen. Genet.** 225: 81-93.
- Lüehrsen, K.R. and Walbot, V. (1994). Addition of A- and U-rich sequence increases the splicing efficiency of a deleted form of a maize intron. **Plant Mol. Biol.** 24: 449-463.
- Lühs, W. and Friedt, W. (1994). The major oil crops. In: Murphy, D.J. (Ed), **Designer Oil Crops Breeding, Processing and Biotechnology** (pp. 27-36). VCH, Weinheim/New York.
- Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987). Selection of AUG codons differs in plants and animals. **EMBO J.** 6: 43-48.
- Maas, C., Laufs, J., Grant, S., Korfhage, C. and Werr, W. (1991). The combination of a novel stimulatory element in the first exon of the maize shrunken-1 gene with the following intron enhances reporter gene expression up to 1000-fold. **Plant Mol. Biol.** 16: 199-207.

- Macdonald, P. and Struhl, G. (1988). *Cis*-acting sequences responsible for anterior localisation of bicoid mRNA in *Drosophila* embryos. **Nature** 336: 596-598.
- Maniatis, T., Fritsch, E.F., Lauer, J. and Lawn, R.M. (1980). The molecular genetics of human hemoglobin. **Annu. Rev. Genet.** 14: 145-178.
- Matzke, M.A. and Matzke, A.J.M. (1995). How and why do plants inactivate homologous (*trans*)genes? **Plant Physiol.** 107: 679-685.
- McConn, M. and Browse, J. (1996). The critical requirement for linolenic acid is pollen development, not photosynthesis, in an *Arabidopsis* mutant. **The Plant Cell** 8: 403-416.
- McDade, L.A. (1990). Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. **Evolution** 44: 1685-1700.
- McDade, L.A. (1992). Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. **Evolution** 46: 1329-1346.
- McKeon, T.A. and Stumpf, P.K. (1982). Purification and characterisation of the stearyl-acyl carrier protein desaturase and the acyl-acyl carrier protein thioesterase from maturing seeds of safflower. **J. Biol. Chem.** 257: 12141-12147.
- Messing, J., Geraghty, D., Herdecker, G., Nien-Tai, H., Kridl, J. and Rubenstein, I. (1983). Plant gene structure. In: Kosuge, T., Merridith, O. and Hollaender, A. (Eds.), **Genetic Engineering in Plants** (pp. 211-227). Plenum Press, New York.
- Meyer, P. and Saedler, H. (1996). Homology-dependent gene silencing in plants. **Annu. Rev. Plant Physiol. Plant Mol. Biol.** 47: 23-48.
- Millar, A.A. and Kunst, L. (1995). Isolation of an *Arabidopsis* cDNA encoding 3-ketoacyl-acyl carrier protein synthase I.1 (GenBank U24177). **Plant Physiol.** 108: 1747.
- Miller, J.F., Zimmerman, D.C. and Vick, B.A. (1987). Genetic control of high oleic acid content in sunflower oil. **Crop Sci.** 27: 923-926.
- Miquel, M. and Browse, J. (1992). *Arabidopsis* mutants deficient in polyunsaturated fatty acid synthesis: Biochemical and genetic characterisation of a plant oleoyl-phosphatidylcholine desaturase. **J. Biol. Chem.** 267: 1502-1509.

- Miquel, M., James Jr., D., Donner, H. and Browse, J. (1993). *Arabidopsis* requires polyunsaturated lipids for low-temperature survival. **Proc. Natl. Acad. Sci. USA** 90: 6208-6212.
- Miquel, M.F. and Browse, J.A. (1994). High-oleate oilseeds fail to develop at low temperature. **Plant Physiol.** 106: 421-427.
- Moustaid, N., Beyers, R.S. and Sul, S.H. (1994). Identification of an insulin response element in the fatty acid synthase promoter. **J. Biol. Chem.** 269: 5629-5634.
- Murphy, D.J. (1992). Modifying oilseed crops for non-edible products. **Trends in Biotechnology** 10: 84-87.
- Murphy, D.J. (1996). Engineering oil production in rapeseed and other oil crops. **Trends in Biotechnology** 14: 206-213.
- Murray, E.E., Rocheleau, T., Everle, M., Stock, C., Sekar, V. and Adang, M. (1991). Analysis of unstable RNA transcripts of insecticidal crystal protein genes of *Bacillus thuringiensis* in transgenic plants and electroporated protoplasts. **Plant Mol. Biol.** 16: 1035-1050.
- Nagai, J. and Bloch, K. (1966). Enzymatic desaturation of stearyl acyl carrier protein. **J. Biol. Chem.** 241: 1925-1927.
- Nagai, J. and Bloch, K. (1968). Enzymatic desaturation of stearyl acyl carrier protein. **J. Biol. Chem.** 243: 4626-4633.
- Nakamura, Y. and Yamada, M. (1974). Long chain fatty acid synthesis in developing castor bean. **Plant Cell Physiol.** 15: 37.
- Nei, M. (1987). **Molecular Evolutionary Genetics**. New York, Columbia University Press.
- Neidleman, S.L. (1987). Effects of temperature on lipid unsaturation. **Biotechnol. Genet. Eng. Rev.** 5: 245-268.
- Nichols, B.W. (1969). The separation, structure and metabolism of monogalactosyl diglyceride species in *Chlorella vulgaris*. **Lipids** 3: 354-360.
- Nichols, B.W., James, A.T. and Breuer, J. (1967). Interrelationships between fatty acid biosynthesis and acyl-lipid synthesis in *Chlorella vulgaris*. **Biochem. J.** 104: 486-496.

- Nishida, I., Beppu, T., Matsuo, T. and Murata, N. (1992). Nucleotide sequence of a cDNA clone encoding a precursor to stearyl-(acyl-carrier-protein) desaturase from spinach, *Spinacia oleracea*. **Plant Mol. Biol.** 19: 711-713.
- Nishida, I., Tasaka, Y., Shiraishi, H. and Murata, N. (1993). The gene and the RNA for the precursor to the plastid-located glycerol-3-phosphate acyltransferase of *Arabidopsis thaliana*. **Plant Mol. Biol.** 21: 267-277.
- Nishiuchi, T., Nakamura, T., Abe, T., Kodama, H., Nishimura, M. and Iba, K. (1995). Tissue-specific and light-responsive regulation of the promoter region of the *Arabidopsis thaliana* chloroplast ω -3 fatty acid desaturase gene (FAD7). **Plant Mol. Biol.** 29: 599-609.
- Nordlund, P., Sjöberg, B.-M. and Eklund, H. (1993). Three-dimensional structure of the free radical protein of ribonucleotide reductase. **Nature** 345: 592-598.
- Norris, S.R., Meyer, S.E. and Callis, J. (1993) The intron of *Arabidopsis thaliana* polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. **Plant Mol. Biol.** 21: 895-906.
- Oard, J.H., Paige, D. and Dvorak, J. (1989). Chimeric gene expression using maize intron in cultured cells of breadwheat. **Plant Cell Rep.** 8: 156-160.
- Ochman, H., Gerber, A.S. and Hartl, D.L. (1988). Genetic applications of an inverse polymerase chain reaction. **Genetics** 120: 621-625.
- Ohlrogge, J.B. (1994). Design of new plant products: engineering of fatty acid metabolism. **Plant Physiol.** 104: 821-826.
- Ohlrogge, J.B. and Browse, J. (1995). Lipid biosynthesis. **The Plant Cell** 7: 957-970.
- Ohlrogge, J.B. and Jaworski, J.G. (1997). Regulation of fatty acid synthesis. **Annu. Rev. Plant Physiol. Plant Mol. Biol.** 48: 109-136.
- Ohlrogge, J.B., Browse, J. and Somerville, C.R. (1991). The genetics of plant lipids. **Biochimica et Biophysica Acta** 1082: 1-26.
- Ohlrogge, J.B., Jaworski, J.G. and Post-Beittenmiller, M.A. (1993). *De novo* fatty acid synthesis. In: Moore, J. (Ed.), **Lipid Metabolism in Plants** (pp. 91-126). CRC Press

- Ohta, T. and Kimura, M. (1973). The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. **Genet. Res.** 22: 201-204.
- Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E. and Browse, J. (1994). *Arabidopsis* FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. **The Plant Cell** 6: 147-158.
- Oskouian, B., Rangan, V.S. and Smith, S. (1997). Regulation elements in the first intron of rat fatty acid synthase gene. **Biochem. J.** 324: 113-121.
- Ottavio, L., Chang, D.C., Rizzo, M.G., Travali, S., Cassadevall, C. and Baserga, R. (1990). Importance of introns in the growth regulation of mRNA levels of the proliferation cell nuclear antigen gene. **Mol. Cell Biol.** 10: 303-309.
- Palmiter, R.D., Dandgren, E.P., Avarbock, M.R., Allen, D.D. and Brinster, R.L. (1991). Heterologous introns can enhance expression of transgenes in mice. **Proc. Natl. Acad. Sci. USA** 88: 478-482.
- Pamilo, P. and Nei, M. (1991). Relationships between gene trees and species trees. **Molec. Biol. Evol.** 5: 568-583.
- Pandey, S.N. and Thejappa, N. (1975). Study on relationship between oil, protein, and gossypol in cottonseed kernels. **JAOCs** 52: 312-315.
- Pandey, S.S. and Subrahmanyam, V.V.R. (1988). Lipid changes in maturing and germinating cottonseeds. **Phytochemistry** 27: 3405-3409.
- Paterson, A.H., Brubaker, C.L. and Wendel, J.F. (1993). A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. **Plant Mol. Biol. Rep.** 11: 122-127.
- Pesole, G., Liuni, S., Grillo, G. and Saccone, C. (1997). Structural and composition features of untranslated region of eukaryotic mRNAs. **Gene** 205: 95-102.
- Phelps, R.A., Cherry, J.P., Hyer, A.H., Cavanaugh, G.C., Harper, G.A., Jernigan, J.E., Hess, D.C. and Lorange, D.G. (1979). Advancements in cottonseed quality-industry's viewpoint. **Cott. Gin Oil Mill Press** 80: 14-15.
- Phelps, R.A., Shenstone, F.S., Kemmerer, A.R. and Evans, R.J. (1964). A review of cyclopropanoid compounds: biological effects of some derivatives. **Poultry Sci.** 44: 358-394.

- Pickett, F.B. and Meeke-Wagner, D.R. (1995). Seeing double: appreciating genetic redundancy. **The Plant Cell** 7: 1347-1356.
- Podkowinski, J., Sroga, G.E., Haselkorn, R. and Gornicki, P. (1996). Structure of a gene encoding a cytosolic acetyl-CoA carboxylase of hexaploid wheat. **Proc. Natl. Acad. Sci. USA** 93: 1870-1874.
- Polashock, J.J., Chin, C.K. and Martin, C.E. (1992). Expression of the yeast Δ -9 fatty acid desaturase in *Nicotiana tabacum*. **Plant Physiol.** 100: 894-901.
- Pope, O.A. and Ware, J.O. (1945). Effects of variety, location, and season on oil, protein, and fuzz of cottonseed and on fibre problems of lint. **USDA Tech. Bull.** 903.
- Post-Beittenmiller, D., Roughan, P.G. and Ohlrogge, J.B. (1992). Regulation of plant fatty acid biosynthesis: analysis of acyl-CoA and acyl-ACP substrate pools in spinach and pea chloroplasts. **Plant Physiol.** 100: 923-930.
- Post-Beittenmiller, M.A., Hlousek-Radojcic, A. and Ohlrogge, J.B. (1989). DNA sequences of a genomic clone encoding an *Arabidopsis* acyl carrier protein (ACP). **Nucl. Acids Res.** 17: 1777.
- Pripbuus, C., Perdereau, D., Fougelle, F., Maury, J., Ferre, P. and Girard, J. (1995). Induction of fatty-acid-synthase gene expression by glucose in primary culture of rat hepatocytes: dependency upon glucokinase activity. **Eur. J. Biochem.** 230: 309-315.
- Purdy, R.H.J. (1986). High oleic sunflower, physical and chemical characteristics. **JAOCS** 63: 1062-1066.
- Purugganan, M. and Wessler, S. (1993). The splicing of transposable elements and its role in intron evolution. In: MacDonald, J.F. (Ed.), **Transposable Elements and Evolution** (pp. 28-36). Kluwer Academic, Dordrecht.
- Pysh, L.D., Aukerman, M.J. and Schmidt, R.J. (1993). OHP1: A maize basic domain/leucine zipper protein that interacts with opaque 2. **The Plant Cell** 5: 227-236.
- Ratray, J. (1991). Plant biotechnology and the oils and fats industry. In: Ratray, J. (Ed.), **Biotechnology of Plant Fats and Oils** (pp. 1-35). Am. Oil Chem. Soc., Champaign, IL.

- Reith, M. (1993). A beta-ketoacyl-acyl carrier protein synthase III gene (*fabH*) is encoded on the chloroplast genome of the red alga *Porphyra umbilicalis*. **Plant Mol. Biol.** 21: 185-189.
- Reith, M. and Munholland, J. (1995). Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. **Plant Mol. Biol. Rep.** 13: 333-335.
- Resenberg, U.B., Preiss, A., Seifert, E., Jackle, H. and Knipple, D.C. (1985). Production of phenocopies by Kruppel antisense RNA injection into *Drosophila* embryos. **Nature** 313: 703-705.
- Rieseberg, L.H. and Soltis, D.E. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. **Evol. Trends Pl.** 5: 65-84.
- Röbbelen, G., Downey, R.K. and Ashri, A. (1989). **Oil Crops of the World**. McGraw-Hill, New York.
- Roder, K., Klein, H., Kranz, H., Beck, K.F. and Schweizer, M. (1994). The tripartite DNA element responsible for diet-induced rat fatty acids synthase (FAS) regulation. **Gene** 144: 189-195.
- Roesler, K.R., Savage, L.J., Shintani, D.K., Shorrosh, B.S. and Ohlrogge, J.B. (1996). Co-purification, co-immunoprecipitation, and co-ordinate expression of acetyl-coenzyme A carboxylase activity, biotin carboxylase, and biotin carboxyl carrier protein of higher plants. **Planta** 198: 517-525.
- Roesler, K.R., Shorrosh, B.S. and Ohlrogge, J.B. (1994). Structure and expression of an *Arabidopsis* acetyl-coenzyme A carboxylase gene. **Plant Physiol.** 105: 611-617.
- Roessler, P.G. and Ohlrogge, J.B. (1993). Cloning and characterisation of the gene that encodes acetyl-coenzyme A carboxylase in the alga *Cyclotella cryptica*. **J. Biol. Chem.** 268: 19254-19259.
- Rosenzweig, A.C., Frederick, C.A., Lippard, S.J. and Nordlund, P. (1993). Crystal structure of a bacterial non-haem iron hydroxylase that catalyses the biological oxidation of methane. **Nature** 366: 537-540.
- Safford, R., Windust, J.H.C., Lucas, C., de Silva, J., James, C.M., Hellyer, A., Smith, C.G., Slabas, A.R. and Hughes, S.G. (1988). Plastid-localised seed acyl-carrier protein of *Brassica napus* is encoded by a distinct, nuclear multigene family. **Eur. J. Biochem.** 174: 287-295.

- Saitou, N. and Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. **Mol. Biol. Evol.** 4: 406-425.
- Sakamoto, T., Wada, H., Nishida, I., Ohmori, M. and Murata, N. (1994). Identification of conserved domains in the delta-12 desaturases of cyanobacteria. **Plant Mol. Biol.** 24: 643-650.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). **Molecular Cloning: A Laboratory Manual** (2nd ed.). Cold Spring Harbour Laboratory, Cold Spring Harbour, NY.
- Sato, A., Becker, C.K. and Knauf, V.C. (1992). Nucleotide sequence of a complementary DNA clone encoding stearyl-acyl carrier protein desaturase from *Simmondsia chinensis*. **Plant Physiol.** 99: 362-363.
- Sayanova, O., Smith, M.A., Lapinskas, P., Stobart, A.K., Dobson, G., Christie, W.W., Shewry, P.R. and Napier, J.A. (1997). Expression of a borage desaturase cDNA containing a N-terminal cytochrome b5 domain results in the accumulation of high levels of Δ^6 -desaturated fatty acids in transgenic tobacco. **Proc. Natl. Acad. Sci. USA** 94: 4211-4216.
- Scheffler, J.A., Sharpe, A.G., Schmidt, H., Sperling, P., Parkin, I.A.P., Lühs, W., Lydiate, D.J. and Heinz, E. (1997). Desaturase multigene families of *Brassica napus* arose through genome duplication. **Theor. Appl. Genet.** 94: 583-591.
- Scherer, D.E. and Knauf, V.C. (1987). Isolation of cDNA clone for the acyl carrier protein-I of spinach. **Plant Mol. Biol.** 9: 127-134.
- Schindler, U., Menkens, A.E., Beckmann, H., Ecker, J.R. and Cashmore, A.R. (1992). Heterodimerization between light-regulated and ubiquitously expressed *Arabidopsis* GBF bZIP proteins. **EMBO J.** 11: 1261-1274.
- Schlötterer, C. and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. **Nucl. Acids Res.** 20: 211-215.
- Schmidt, H., Dresselhaus, T., Buck, F. and Heinz, E. (1994). Purification and PCR-based cDNA cloning of a plastidial n-6 desaturase. **Plant Mol. Biol.** 26: 631-642.
- Schuch, R., Brummel, M. and Spener, F. (1993). Medium-chain acyl-ACP thioesterase is not the exclusive enzyme responsible for early chain-length termination in medium-chain fatty acid synthesis. In: **International Workshop on "Biotechnological Aspects of Plant Lipids"**, Carmona, Spain.

- Schulte, W., Töpfer, R., Stracke, R., Schell, J. and Martini, N. (1997). Multi-functional acetyl-CoA carboxylase from *Brassica napus* is encoded by a multi-gene family: indication for plastidic localisation of at least one isoform. **Proc. Natl. Acad. Sci. USA** 94: 3465-3470.
- Schultz, D.J., Cahoon, E.B., Shanklin, J., Craig, R., Cox-Foster, D.L., Mumma, R.O. and Medford, J.I. (1996). Expression of a $\Delta 9$ 14:0-acyl carrier protein fatty acid desaturase gene is necessary for the production of $\omega 5$ anacardic acids found in pest-resistant geranium (*Pelargonium xhortorum*). **Proc. Natl. Acad. Sci. USA** 93: 8771-8775.
- Schultz, J.R., Tansey, T., Gremke, L. and Storti, R.V. (1991). A muscle-specific enhancer required for rescue of indirect flight muscle and jump muscle function regulates *Drosophila* tropomyosin I gene expression. **Mol. Cell Biol.** 11: 1901-1911.
- Seelanan, T., Schnabel, A. and Wendel, J.F. (1997). Congruence and consensus in the cotton tribe. **Syst. Bot.** 22: 259-290.
- Seiler, G.J. (1994). Oil concentration and fatty acid composition of Achenes of North American *Helianthus* (Asteraceae) species. **Economic Botany** 48: 271-279.
- Shah, F.H. and Rashid, O. (1996). Nucleotide sequence of a complementary DNA clone encoding stearyl-acyl-carrier-protein (accession No. U68756) from *Elaeis guineensis* var. tenera. **Plant Physiol.** 112: 1399.
- Shanklin, J. and Somerville, C. (1991). Stearyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. **Proc. Natl. Acad. Sci. USA** 88: 2510-2514.
- Shanklin, J., Mullins, C. and Somerville, C. (1991). Sequence of a complementary DNA from *Cucumis sativus* L. encoding the stearyl-acyl-carrier-protein desaturase. **Plant Physiol.** 97: 467-468.
- Shanklin, J., Whittle, E. and Fox, B.G. (1994). Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. **Biochemistry** 33: 12787-12794.
- Shimakata, T. and Stumpf, P.K. (1982). Isolation and function of spinach leaf β -ketoacyl-(acyl-carrier-protein) synthases. **Proc. Natl. Acad. Sci. USA** 79: 5808-5812.

- Shorrosh, B.S., Dixon, R.A. and Ohlrogge, J.B. (1994). Molecular cloning, characterisation, and elicitation of acetyl-CoA carboxylase from alfalfa. **Proc. Natl. Acad. Sci. USA** 91: 4323-4327.
- Shorrosh, B.S., Roesler, K.R., Shintani, D.K., van der Loo, F.J. and Ohlrogge, J.B. (1995). Structural analysis, plastid localisation, and expression of the biotin carboxylase subunit of acetyl-Coenzyme A carboxylase from tobacco. **Plant Physiol.** 108: 805-812.
- Siggaard-Andersen, M., Kauppinen, S. and von Wettstein-Knowles, P. (1991). Primary structure of a cerulenin-binding β -ketoacyl-[acyl carrier protein] synthase from barley chloroplasts. **Proc. Natl. Acad. Sci. USA** 88: 4114-4118.
- Simpson, C.G. and Brown, J.W.S. (1993). Efficient splicing of an AU-rich antisense intron sequence. **Plant Mol. Biol.** 21: 205-211.
- Simpson, C.G. and Filipowicz, W. (1996). Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery. **Plant Mol. Biol.** 32: 1-41.
- Simpson, C.G., Leader, D.J. and Brown, J.W.S. (1993). **Plant Molecular Biology Labfax**. BIOS Scientific Publishers, Oxford.
- Singh, S., McKinney, S. and Green, A. (1994). Sequences of a cDNA from *Linum usitatissimum* encoding the stearyl-Acyl carrier protein desaturase. **Plant Physiol.** 104: 1075.
- Singh, S.P., Shukla, S., Khanna, K.R. and Banerji, R. (1998). Variation of major fatty acids in F₈ generation of opium poppy (*Papaver somniferum* x *Papaver setgerum*) genotypes. **J. Sci. Food Agri.** 76: 168-172.
- Singh, S.P., van der Heide, T., McKinney, S. and Green, A.G. (1995). Nucleotide sequence of a cDNA (accession No. X91139) from *Brassica juncea* encoding a microsomal omega-6 desaturase. **Plant Physiol.** 109: 1498.
- Slabas, A.R., Chase, D., Nishida, I., Murata, N., Sidebottom, C., Safford, R., Sheldon, P.S., Kekwick, R.G.O., Hardie, D.G. and Mackintosh, R.W. (1992). Molecular cloning of higher-plant 3-oxoacyl-(acyl carrier protein) reductase -sequence identities with the nodG-gene product of the nitrogen fixing soil bacterium *Rhizobium meliloti*. **Biochem. J.** 283: 321-326.

- Slabaugh, M., Leonard, J., Tai, J., Jaworski, J. and Knapp, S. (1993). Condensing enzymes and thioesterase expressed in immature embryos of *Cuphea wrightii*. In: **The Proceedings of Plant Lipid Symposium**, Minneapolis, Minnesota.
- Slabaugh, M.B., Tai, H., Jaworski, J. and Knapp, S.J. (1995). cDNA clones encoding β -ketoacyl-acyl carrier protein synthase III from *Cuphea wrightii*. **Plant Physiol.** 108: 443-444.
- Slocombe, S.P., Cummins, I., Jarvis, R.P. and Murphy, D.J. (1992). Nucleotide sequence and temporal regulation of a seed-specific *Brassica napus* cDNA encoding a stearyl-acyl carrier protein (ACP) desaturase. **Plant Mol. Biol.** 20: 151-155.
- Slocombe, S.P., Piffanelli, P., Fairbairn, D., Bowra, S., Hatzopoulos, P., Tsiantis, M. and Murphy, D.J. (1994). Temporal and tissue-specific regulation of a *Brassica napus* stearyl-acyl carrier protein desaturase gene. **Plant Physiol.** 104: 1167-1176.
- Smith, M.A., Cross, A.R., Jones, O.T.G., Griffiths, W.T., Stymne, S. and Stobart, K. (1990). Electron-transport components of the 1-acyl-2-oleoyl-sn-glycero-3-phosphocholine Δ^{12} -desaturase (Δ^{12} -desaturase) in microsomal preparations from developing safflower (*Carthamus tinctorius* L.) cotyledons. **Biochem. J.** 272: 23-29.
- Smith, M.A., Stobart, A.K., Shewry, P.R. and Napier, J.A. (1994). Tobacco cytochrome b_5 : cDNA isolation, expression analysis and *in vitro* protein targeting. **Plant Mol. Biol.** 25: 527-537.
- Smouse, T.H. (1979). A review of soybean oil reversion flavour. **JAOCs** 56: 747-751.
- Soltis, D.E. and Soltis, P.S. (1993). Molecular data and the dynamic nature of polyploidy. **Critical Reviews in Plant Sciences** 12: 243-273.
- Sonenberg, N. (1994). mRNA translation: influence of the 5' and 3' untranslated regions. **Curr. Opin. Gen. Dev.** 4: 15-19.
- Souciet, G. and Weil, J.H. (1992). Molecular cloning and sequencing of a cDNA encoding the maize (*Zea mays* L.) acyl carrier protein (ACP); comparison to other ACP genes and ACPs. **Plant Sci.** 81: 215-225.
- Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. **J. Mol. Biol.** 98: 503-517.

- Spychalla, J.P., Kinney, A.J. and Browse, J. (1997). Identification of an animal ω -3 fatty acid desaturase by heterologous expression in *Arabidopsis*. **Proc. Natl. Acad. Sci. USA** 94: 1142-1147.
- Stansbury, M.F., Cucullu, A.F. and Den Hartog, T.H. (1954). Cottonseed contents variation: influence of variety and environment on oil content of cottonseed kernels. **Agric. Food Chem.** 2: 692-696.
- Stansbury, M.F., Pons, W.A. and Den Hartog, G.T. (1956). Relationship between oil, nitrogen, and gossypol in cottonseed kernels. **JAOCs** 30: 120-123.
- Stanton, J.M. and Blumenfeld, J.K. (1992). Modifying oil yield affects profile too. **INFORM** 3: 1019-1022.
- Stefansson, B.R., Hougen, F.W. and Downey, R.K. (1961). Note on the isolation of rape plants with seed oil free from erucic acid. **Can. J. Plant Sci.** 41: 218-219.
- Stewart, J.M. (1995). Potential for crop improvement with exotic germplasm and genetic engineering. In: Constable, G.A. and Forrester, N.W. (Eds.), **Challenging the Future: Proceedings of the World Cotton Research Conference**, Melbourne, CSIRO (pp. 313-327).
- Stukey, J.E., McDonough, V.M. and Martin, C.E. (1990). The *OLE1* gene of *Saccharomyces cerevisiae* encodes the delta 9 fatty acid desaturase and can be functionally replaced by the rat stearyl-CoA desaturase gene. **J. Biol. Chem.** 265: 20144-20149.
- Stumpf, P.K. (1989). Biosynthesis of fatty acids in higher plants. In: Röbbelen, G., Downey, R.K. and Ashri, A. (Eds.), **Oil Crops in the World** (pp. 38-62). McGraw Hill, New York.
- Stymne, S. and Appelqvist, L.A. (1978). The biosynthesis of linolenate from oleoyl-CoA via oleoyl-phosphatidylcholine in microsomes of developing safflower seeds. **Eur. J. Biochem.** 90: 223-247.
- Stymne, S. and Appelqvist, L.A. (1980). Biosynthesis of linoleate and linolenate in homogenates from developing soya bean cotyledons. **Plant Sci. Lett.** 17: 287-294.
- Stymne, S. and Stobart, A.K. (1987). Triacylglycerol biosynthesis. In: Stumpf, P.K. (Ed.), **The Biochemistry of Plants: A Comprehensive Treatise** (pp. 175-214). Academic Press, New York.

- Swofford, D.L. (1990). **PAUP: Phylogenetic Analysis Using Parsimony**, version 3.1.1. Illinois Natural History Survey, Champaign, IL.
- Tai, H. and Jaworski, J.G. (1993). 3-ketoacyl-acyl carrier protein synthase III from spinach (*Spinacia oleracea*) is not similar to other condensing enzymes of fatty acid synthase. **Plant Physiol.** 103: 1361-1367.
- Tai, H., Post-Beittenmiller, D. and Jaworski, J.G. (1994). Cloning of a cDNA encoding 3-ketoacyl-acyl carrier protein synthase III from *Arabidopsis*. **Plant Physiol.** 106: 801-802.
- Tauz, D. (1989). Hepervariability of simple sequences as a general source for polymorphic DNA markers. **Nucl. Acids Res.** 17: 6463-6471.
- Taylor, M.A., Smith, S.B., Davies, H.V. and Burch, L.R. (1992). The primary structure of a cDNA clone of the stearyl-acyl carrier protein desaturase gene from potato (*Solanum tuberosum* L.). **Plant Physiol.** 100: 533-534.
- Thiede, M.A., Ozoles, J. and Strittmatter, P. (1985). The induction and characterisation of cDNA for rat liver stearyl-CoA desaturase mRNA. **J. Biol. Chem.** 260: 14459-14463.
- Thiede, M.A., Ozols, J. and Strittmatter, P. (1986). Construction and sequence of cDNA for rat liver stearyl coenzyme A desaturase. **J. Biochem.** 261: 13230-13235.
- Thomas, T.L. (1993). Gene expression during plant embryogenesis and germination: an overview. **The Plant Cell** 5: 1401-1410.
- Thomas, Y.L., Vivekananda, J. and Bogue, M.A. (1991). ABA regulation of gene expression in embryos and mature plants. In: Davies, W.J. and Jones, H.G. (Eds.), **Abscisic Acid: Physiology and Biochemistry** (pp. 125-135). Bios Scientific Publishers, Oxford.
- Thompson, G.A., Scherer, D.E., Aken, S.F., Kenny, J.W., Young, H.L., Shintani, D.K., Kridl, J.C. and Knauf, V.C. (1991). Primary structures of the precursor and mature forms of stearyl-acyl carrier protein desaturase from safflower embryos and requirement of ferredoxin for enzyme activity. **Proc. Natl. Acad. Sci. USA** 88: 2578-2582.

- Thompson, G.A.J. (1993). Response of lipid metabolism to developmental change and environmental perturbation. **Lipid Metabolism in Plants**. Boca Raton, FL, CRC Press. 591-619.
- Töpfer, R., Martini, N. and Schell, J. (1995). Modification of plant lipid synthesis. **Science** 268: 681-686.
- Touma-Touchan, H. (1977). Etudes biochimiques et ultrastructurales des lipides dans la graine du cotonnier. **J. Ultrastruct. Res.** 58: 271-288.
- Trizzini, G.F., Horrichs, A. and Somssich, I.E. (1993). Isolation of putative defense-related genes from *Arabidopsis thaliana* and expression in fungal elicitor-treated cells. **Plant Mol. Biol.** 21: 385-389.
- Trueba, G.A. and Johnson, R.C. (1996). Random primed gene walking PCR: a simple procedure to retrieve nucleotide fragments adjacent to known DNA sequences. **BioTechniques** 21: 20.
- Uchimiya, H., Kidou, S., Shimazaki, T., Aotsuka, S., Takamatsu, S., Nisi, R., Hashimoto, H., Matasubayashi, Y., Kidou, N., Umeda, M. and Kato, A. (1992). Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.). **The Plant Journal** 2: 1005-1009.
- van der Krol, A.R., Lenting, P.E., Veenstra, J., van der Meer, I.M., Koes, R.E., Gerats, A.G.M., Mol, J.N.M. and Stuitje, A.R. (1988). An anti-sense chalcone synthase gene in transgenic plants inhibits flower pigmentation. **Nature** 333: 866-869.
- van der Loo, F.J. and Somerville, C. (1994). Plant ω -3 fatty acid desaturase cDNA from *Ricinus communis*. **Plant Physiol.** 105: 443-444.
- van der Loo, F.J., Brou, P., Turner, S. and Somerville, C.R. (1995). An oleate 12-hydroxylase from castor (*Ricinus communis* L.) is a fatty acyl desaturase homolog. **Proc. Natl. Acad. Sci. USA** 92: 6743-6747.
- Vasil, V., Clancy, M., Ferl, R.J. and Vasil, I.K. (1989). Increased gene expression by the first intron of maize shrunken-1 locus in grass species. **Plant Physiol.** 91: 1575-1579.
- Voelker, T.A. (1995). Genetic manipulation of plant oil composition. In: **Induced Mutations and Molecular Techniques for Crop Improvement** (pp. 93-99). IAEA and FAO, Vienna.

- Voelker, T.A., Worrell, A.C., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D.J., Radke, S.E. and Davies, H.M. (1992). Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. **Science** 257: 72-74.
- Voetz, M., Klein, B., Schell, J. and Töpfer, R. (1994). Three different cDNAs encoding acyl carrier proteins from *Cuphea lanceolata*. **Plant Physiol.** 106: 785-786.
- Walsh, M.C., Klopfenstein, W.E. and Harwood, J.L. (1990). The short chain condensing enzyme has a widespread occurrence in the fatty acid synthetases from higher plants. **Phytochemistry** 29: 3797-3799.
- Weaire, P.J. and Kekwick, R.G.O. (1975). Synthesis of fatty acids in avocado mesocarp and cauliflower bud tissue. **Biochem. J.** 146: 425-432.
- Weber, K. and Kabsch, W. (1994). Intron positions in actin genes seem unrelated to the secondary structure of the protein. **EMBO J.** 13: 1280-1286.
- Weinreb, A., Collier D.A., Birshtein, B.K. and Wells, R.D. (1990). Left-handed Z-DNA and intramolecular triplex formation at the site of an unequal sister chromatid exchange. **J. Biol. Chem.** 265: 1352-1359.
- Wells, R.D., Collier, D.A., Hanvey, J.C., Shimizu, M. and Wohlrab, F. (1988). The chemistry and biology of unusual DNA structures adopted by oligopurine oligopyrimidine sequence. **FASEB J.** 2: 2939-2949.
- Wendel, J.F. and Albert, V.A. (1992). Phylogenetics of the cotton genus (*Gossypium*): Character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. **Syst. Bot.** 17: 115-143.
- Wendel, J.F., Rowley, R. and Stewart, J. (1994). Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (*Malvaceae*). **Plant Syst. Evol.** 192: 49-59.
- Wendel, J.F., Schnabel, A. and Seelanan (1995a). An unusual ribosomal DNA sequence form *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. **Mol. Phyl. Evol.** 4: 298-313.
- Wendel, J.F., Schnabel, A. and Seelanan, T. (1995b). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). **Proc. Natl. Acad. Sci. USA** 92: 280-284.

- Wendel, J.F., Stewart, J.McD. and Rettig, J.H. (1991). Molecular evidence for homoploid reticulate evolution among Australian species of *Gossypium*. **Evolution** 45: 694-711.
- White, A.J., Elborough, K.M., Jones, H. and Slabas, A.R. (1996). Transgenic modification of acetyl CoA carboxylase and β -keto reductase levels in *Brassica napus* functional and regulatory analysis. In: Williams, J.P., Khan, M.U. and Lem, N.W. (Eds.), **Physiology, Biochemistry and Molecular Biology of Plant Lipids** (pp. 14-16). Kluwer Academic Publishers, Dordrecht/Boston/London.
- Williams, J.P., Francisco, O. and Khan, M.U. (1997). A low temperature-sensitive mutant of *Brassica napus* that reduces the level of trienoic fatty acid in chloroplast lipids. **Plant Physiology and Biochemistry** 35: 257-264.
- Wissenbach, M., Siggaard-Andersen, M., Kauppinen, S. and von Wettstein-Knowles, P. (1992). Condensing enzymes of barley. In: Cherif, A., Ben Miles-Daoud, D., Marzouk, B., Smaoui, A. and Zarouk, M. (Eds.), **Metabolism Structure and Utilisation of Plant Lipids** (pp. 393-396). Centre National Pédagogique, Tunis.
- Wolter, F.P. (1996). Conserved intron position in 3' untranslated region of a cDNA encoding the plastidial sn-glycerol-3-phosphate acyltransferase of spinach (accession No. Z49091). **Plant Physiol.** 112: 1735.
- Yadav, N.S., Wierzbicki, A., Aegerter, M., Caster, C.S., Perez-Grau, L., Kinney, A.J., Hiltz, W.D., Booth Jr, R., Schweiger, B., Stecca, K.L., Allen, S.M., Blackwell, M., Reiter, R.S., Carlson, T.J., Russell, S.H., Feldmann, K.A., Pierce, J. and Browse, J (1993). Cloning of higher plant ω -3 fatty acid desaturases. **Plant Physiol.** 103: 467-476.
- Yamamoto, K.T. (1994). Further characterisation of auxin-regulated mRNAs in hypocotyl sections of mung bean [*Vigna radiata* (L.) Wikcozek]: sequence homology to genes for fatty-acid desaturases and a typical late-embryogenesis-abundant protein, and the mode of expression of the mRNAs. **Planta** 192: 359-364.
- Yuan, L., Voelker, T.A. and Hawkins, D.J. (1995). Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. **Proc. Natl. Acad. Sci. USA** 92: 10639-10643.
- Yukawa, Y., Takaiwa, F., Shoji, K., Masuda, K. and Yamada, K. (1996). Structure and expression of two seed-specific cDNA clones encoding stearyl-acyl carrier

protein desaturase from Sesame, *Sesamum indicum* L. **Plant Cell Physiol.** 37: 201-205.

- Zhang, S.H., Lawton, M.A., Hunter, T. and Lamb, C.J. (1994). atpk1, a novel ribosomal protein kinase gene from *Arabidopsis*. I. Isolation, characterisation, and expression. **J. Biol. Chem.** 269: 17586-17592.
- Zhao, X., Wing, R.A. and Paterson, A.H. (1996). Cloning and characterisation of the majority of repetitive DNA in cotton (*Gossypium* L.). **Genome** 38: 1177-1188.
- Zhu, Q. (1996). RNA polymerase II-dependent plant *in vitro* transcription systems. **The Plant Journal** 10: 185-188.
- Zou, J.T., Abrams, G.D., Barton, D.L., Taylor, D.C., Pomeroy, M.K. and Abrams, S.R. (1995). Induction of lipid and oleosin biosynthesis by (+)-abscisic acid and its metabolites in microspore-derived embryos of *Brassica napus* L. cv Reston. **Plant Physiol.** 108: 563-571.