Identification and characterisation of cotton boll wall-specific promoters

Damien James Lightfoot, B. Sc. (Hons.)

A thesis submitted for the degree of Doctor of Philosophy

Discipline of Genetics
School of Molecular and Biomedical Science
The University of Adelaide
November 2006

Table of contents

Index of figures and tables

Corrections

Page 4, section 1.4.3, line 7: delete "are these".

Page 9, line 11: replace "resistance to the endotoxin" with "genetic resistance to anti-pest treatments".

Page 11, section 1.8, second paragraph, line 13: replace "Arial" with "Aerial".

Page 11, section 1.8, second paragraph, line 9: delete "of".

Page 14, line 3: replace "1997" with "1996".

Page 15, line 1: replace "pupae busting" with "the destruction of pupal burrows".

Page 22, paragraph 3, line 12: replace "While the commercialised variety was successful in preventing insect attack, the high level of transgenic protein in the pollen may have been harmful to the monarch butterfly (*Danaus plexippus*) and the variety was withdrawn from the market." with "While the commercialised variety was successful in preventing insect attack, the high level of transgenic protein in the pollen may have been harmful to the monarch butterfly (*Danaus plexippus*). The variety was withdrawn from the market due to these non-target effects as well as concerns about patent and performance issues."

Page 25, section 2.1.1: replace whole section with:

2.1.1: Plant material

All cotton material was isolated from plants grown from seed provided by either Cotton Seed Distributors (Narrabri, NSW) or the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) Cotton Research Unit (Narrabri, NSW) as indicated below. Cotton species and varieties used were:

Cotton Seed Distributors:
G. hirsutum cultivar Siokra 1-4
G. hirsutum cultivar Sicot 189
G. hirsutum cultivar Sicot V-2

CSIRO:
G. raimondii
G. herbaceum variety *africanum*

Page 54, figure legend: replace whole legend with:

Figure 3.7: Northern analysis of boll wall-preferential mRNAs.
The abundance of the six mRNA transcripts in various tissues was assessed by Northern analysis. The tissues used were 0, 5, 10, 15, 20, 25, 30, 40 and 50 DPA boll wall, young leaf, 3 to 6 DPA fibre, white film from 5 to 10 DPA bolls, 0 to 5 DPA bract, 5 to 10 DPA calyx, 0 DPA petal and 0 DPA staminal column. Root, stem and cotyledon tissues were obtained from seedlings of approximately seven centimetres in height. 10μg of each RNA was electrophoresed, transferred to a membrane and hybridised with each of the cDNAs. The bottom panel shows a representative ethidium bromide stained RNA gel used to monitor RNA loading. Marker lane sizes are indicated in nucleotides.

Page 62, second paragraph, line 24: replace "ball" with "boll".

Page 72, second paragraph line 3: replace "The high abundance of SuSy mRNA in the boll wall corresponds with the timing of growth and secondary cell wall synthesis of the cotton fibres." with "The peak abundance of SuSy mRNA in the boll wall corresponds with the growth phase of the fibres, with lower abundances of RNA present during secondary cell wall synthesis."

Page 92, line 1: replace "none" with "non".

Page 106, line 4: replace "cotton fibres" with "cotton ovules and attached fibres".

Pages 165 – 177: italicise all occurrences of "*Bacillus thuringiensis*", "*Trifolium subterraneum*", "*Bt*", "*Viola hederacea*", "*cry1Ab*", "*Gossypium hirsutum*", "*Heliothis*", "*Vitis vinifera*", "*Cajanus cajan*", "*Oryza sativa*", "*Medicago sativa*", "*Citrus sinensis*", "*Arabidopsis thaliana*", "*CaPRP1*", "*Capsicum annuum*", "*Populus tremuloides*", "*Tetrastichus howardi*", "*Secale cereale*", "*Solanum tuberosum*", "*Manihot esculenta*", "*Nicotiana glauca*", "*Camellia sinensis*", "*Manduca* sexta", "*Spodoptera littoralis*", "*Helicoverpa armigera*" and "*Lumbricus terrestris*".

Page 167, References. Insert the following:

Clark, B. W., T. A. Phillips and J. R. Coats. 2005. Environmental fate and effects of *Bacillus thuringiensis* proteins from transgenic crops: a review. Journal of Agricultural and Food Chemistry 53: 4643-4653.

Edge, J. M., J. H. Benedict, J. P. Carroll and H. K. Reding. 2001. Contemporary issues. Bollgard cotton: An assessment of global economic, environmental, and social benefits. The Journal of Cotton Science 5: 121-136.

Page 167, line 30: replace "*gossypium*" with "*Gossypium*".

Page 173, References: Insert the following:

Perlak, F. J., R. L. Fuchs, D. A. Dean, S. L. McPherson and D. A. Fischhoff. 1991. Modification of the coding sequence enhances plant expression of insect control protein genes. Proceedings of the National Academy of Sciences of the United States of America 88: 3324-3328.

Declaration

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

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

_____

Damien Lightfoot

Acknowledgements

I would like to thank my principal supervisor Jeremy Timmis for giving me the opportunity to conduct my honours and PhD research in his lab. His guidance and help throughout the last five and a half years have been invaluable. I would also like to thank my co-supervisor Sharon Orford, particularly for her help with all of those complicated cotton methods!

A big thanks also goes to the CRDC (Cotton Research and Development Corporation) who generously supported my project and gave me a great introduction to the Australian cotton industy.

Thanks to past and present members of the Timmis lab who have been a great help to me with my work and who have made the lab a truly great place to work. Special thanks to Chunyuan Huang for all of his unconventional, but successful, methods, to Sven Delaney for his help and advice, to Sarah Harmer for her help with all of the cotton work, and to John Humphries for his help and assistance and for being a great friend in the lab.

I would also like to thank all of the members of the genetics disipline. Thanks to everyone who gave me ideas, let me use their equipment and helped me during my time in the department.

I would like to give a final thanks to everyone from the Molecular Life Sciences Building, past and present, who have made it a fun place to work, especially the regular Friday night Uni Bar crowd.

Abstract

The cotton boll contains the seeds of the plant to which long, white fibres are attached. The cotton industry takes advantage of these fibres to spin yarns for textile production. A major challenge facing the cotton industry is that of crop loss to insect attack. The primary insect pests of cotton preferentially attack the boll, causing damage to the commercially important fibres. The recent introduction of *Bt*-transgenic varieties, containing genes with anti-pest properties from the soil bacterium *Bacillus thuringiensis*, has had positive impacts on pest control and pesticide usage. These transgenes are under control of constitutive promoters, resulting in endotoxin expression in all parts of the plant. This constant high level transgene expression may have several detrimental effects, such as placing strong selective pressure on pest populations to develop resistance, non-target effects of the transgene on other organisms, a yield penalty to the plant, and the presence of transgenic protein in secondary commercial products. For these reasons, this project aims to identify promoters that could be used for tissue-specific expression of anti-pest molecules in only the boll wall of the plant. A differential screening approach was used to identify several boll wall-specific mRNAs and the temporal and spatial abundance of these transcripts was determined using Northern analysis. The promoters corresponding to these transcripts were identified using Genome Walker® PCR and isolated from genomic DNA by PCR. Transient transformation of various cotton tissues with these promoters driving reporter expression resulted in predominant boll wall expression. The cotton promoters identified here provide an alternative tool to constitutive promoters for use in future transgenic varieties.

Chapter 1: Introduction

1.1: The cotton plant and industry

Cotton is grown as an annual crop in more than 70 countries and contributes approximately US$28 billion to the world's economy every year (ICAC, 2005[1]). The primary commercial product of the cotton plant is the seed coat hair or fibres, which are removed from the seed for use in textile production. Secondary commercial products derived from the processed seed include oils and feedstock. In Australia, cotton is grown commercially in Queensland and New South Wales, with an estimated crop value of AU$1.16 billion in 2006 (CRDC, 2006[2]).

1.2: Cotton plant and flower development

Cultivated cotton is usually grown as an annual crop with mature plants reaching a height of approximately 1.5 metres. Flower buds (squares) begin to develop about 45 days after germination and flowering (anthesis) occurs approximately 20 days later. The flowers are surrounded by an epicalyx that consists of three large bracts that remain closed until they are pushed open by the elongating flower bud on the day prior to anthesis. Within the epicalyx is a small cup-shaped calyx, consisting of five short green sepals, which surrounds the lower end of the ovary. The flower consists of five creamy white overlapping petals, fused at their base, that turn pink one day post anthesis (DPA). The cotton flower contains a staminal column with numerous stamens surrounding a style, which leads to an ovary at the flowers base. The ovary has three to five compartments with five to twelve ovules in each compartment (figure 1.1). On the day of anthesis, flowers typically open around dawn, self-pollinate and close their petals within 24 hours. Following anthesis, the bracts close around the ovary (young boll) and the flower structure is shed. The ovary develops into a mature boll over a period of approximately 60 days depending upon conditions such as temperature, light and water availability.

The cotton boll expands rapidly during the first 15 days of its development from ~10 mm at 0 DPA to ~40 mm at 15 DPA (Wilkins and Jernstedt, 1999). This period of boll expansion corresponds with the period of elongation of the fibres contained within the boll. Fibres elongate at a peak rate between approximately 6 and 12 DPA, with elongation ceasing around 20 DPA (Meinert and Delmer, 1977), by which time the boll has reached its maximum size (Matthews and Tunstall, 1994). After 20 DPA, fibre maturation without elongation

[1] International Cotton Advisory Committee (http://www.icac.org)
[2] Cotton Research and Development Corporation (http://www.crdc.com.au)

Figure 1.1: Cross section of a cotton flower.
Cross-sectional representation of a cotton flower at the day of anthesis showing the location of various floral structures. The ovules are contained within the ovary (shaded green), which develops into the boll after anthesis. Adapted from Raven *et al*. (1992).

continues until approximately 60 DPA. Around this time, the boll reaches maturity and splits open along the carpel boundaries, exposing the seeds with attached fibres.

1.3: Cotton genetics and evolution

Cultivated cottons belong to the genus *Gossypium* L. with ~50 species characterised to date (Fryxell, 1992). 45 of these are diploid species (2n=26) that are categorised in eight distinct genomic groups, A-G and K, based on chromosome pairing relationships (Endrizzi *et al.*, 1985). The genomic differences that have evolved between these eight groups are believed to result from geographic isolation of the corresponding diploid species. The C, G and K genomes originated in Australia, the D genome in the Americas and the A, B, E and F genomes in Africa and Asia.

The tetraploid cottons (2n=4x=56), comprised of five species, are thought to have formed through the hybridisation of two diploid species, containing the A and D genomes, in the A genome cytoplasm (Wendel, 1989). This process is thought to have taken place between 1.1 and 1.9 million years ago in Southern or Central America following ocean migration of the A genome (Asia/Africa) to the Americas and subsequent chromosome doubling to produce allotetraploid species (AADD genome) (Wendel, 1989).

Only those species that possess the A genome are capable of producing commercially valuable fibres (Wendel, 1989). Four species are currently commercially cultivated: the AA diploids (*G. herbaceum* and *G. arboreum*) and the AADD tetraploids (*G. hirsutum* and *G. barbadense*). The tetraploid cottons are agronomically superior to the diploid cottons, producing higher yields and longer fibres, and have adapted to a wide range of environments (Horsefall, 1972). The AADD tetraploids are the most widely cultivated, with *G. hirsutum* and *G. barbadense* accounting for 95% and 4%, respectively, of world fibre production (Cedroni *et al.*, 2003).

It is thought that A genome diploids were domesticated around 6000 BC (Moulherat *et al.*, 2002) and AADD tetraploid cotton between 3500 and 2300 BC (Stephens and Moseley, 1974). Since then, selective breeding of *G. hirsutum* has emphasised maximum yield, whereas *G. barbadense* is grown for its superior fibre length, fineness and strength (Jiang *et al.*, 1998). *G. hirsutum* varieties yield fibres of 20 to 25 mm in length, while *G. barbadense* varieties produce longer fibres of greater than 35 mm (Basra and Malik, 1984).

1.4: Cotton crop entomology

1.4.1: Cotton crop stresses

Each growing season, cotton crops endure biotic and abiotic stresses that negatively impact on plant growth, as well as fibre yield and quality. Biotic stresses include pest attack and competition by weeds, while abiotic stresses are usually related to climatic conditions such as water availability and temperature. It is the biotic stresses in the form of pest attack that are of particular interest to this study. Damage caused by cotton pests has the potential to pose a severe problem. Studies indicate that cotton crop yields would be reduced by up to 84% without the use of pesticides to control pests (Oerke *et al.*, 1994).

1.4.2: Global cotton pests

The types of pests that inhabit cotton crops vary both globally and locally. Differences in pest populations are caused by many factors, including pest establishment history, climate, neighbouring crops, pest treatments, local farm practices and geography. It has been estimated that there are up to 60 pests of cotton crops, with the majority of damage caused by five to ten key pests in most production areas (Hearn and Fitt, 1992). Worldwide, the boll weevil (*Anthonomus grandis*), and lepidopteran larvae (primarily *Helicoverpa* and *Pectinophora spp*.) cause the most damage (Matthews and Tunstall, 1994).

1.4.3: Australian cotton entomology

More than 1000 different species of insects and arachnids (spiders and mites) have been observed in Australian cotton fields, with only 30 of these species recorded as causing damage (table 1.1) (Pyke and Brown, 1996). Some species are beneficial to the cotton crop because they feed on pests, but the majority are benign or neutral. Despite the fact that there are 30 species with the potential to cause crop damage, only three pests are of major importance, requiring some management in most regions in most seasons. These three primary pests, in order of importance, are these are two Australian *Helicoverpa* species: the cotton bollworm (*Helicoverpa armigera*) and the native budworm (*Helicoverpa punctigera*) and the two-spotted spider mite (*Tetranychus urticae*). The cotton aphid (*Aphis gossypii*) can also be an important pest that requires treatment, but it is usually kept under control by the management measures used for the three primary pests (Pyke and Brown, 1996).

Australian cotton pests

Lepidopteran insects cause the most damage to Australian cotton crops with *H. armigera* and *H. punctigera* accounting for the vast majority of this damage (Fitt, 1994). The

Species name	Common name	Damage	Damage type					
			Square	Boll	Flower	Term.	Leaf	Seed.
Lepidoptera								
Helicoverpa armigera	Cotton bollworm	1	Y	Y	Y	Y		
Helicoverpa punctigera	Native budworm	1	Y	Y	Y	Y		
Pectinophora scutigera	Pink-spotted bollworm	3	Y	Y	Y	Y		
Pectinophora gossypiella	Pink bollworm	3	Y	Y	Y	Y		
Spodoptera litura	Cluster caterpillar	3			Y		Y	
Anomis flava	Cotton looper	3					Y	
Crocidosema plebejana	Cotton tipworm	3				Y		
Spodoptera exigua	Beet armyworm	3				Y	Y	
Spodoptera frugiperda	Fall armyworm	3				Y	Y	
Earias heugeli	Rough bollworm	3	Y	Y				
Agrotis infusa	Common cutworm	3						Y
Bucculatrix gossypii	Cotton leaf perforator	3					Y	
Mites								
Tetranychus urticae	Two-spotted spider mite	2					Y	
Tetranychus ludeni	Bean spider mite	3					Y	
Tetranychus lambi	Strawberry spider mite	3					Y	
Aphids								
Aphis gossypii	Cotton aphid	2		Y			Y	
Aphis craccivora	Cowpea aphid	3		Y			Y	
Myzus persicae	Green peach aphid	3		Y			Y	
Smynthurodes betae	Bean root aphid	3		Y			Y	
Thrips								
Thrips tabaci	Tobaco thrip	2				Y	Y	
Thrips imaginis	Plague thrip	3				Y	Y	
Frankliniella schultzei	Tomato thrips	3				Y	Y	
Frankliniella occidentalis	Western flower thrip	3				Y	Y	
Mirids								
Creontiades dilutus	Green mirid	2	Y	Y		Y	Y	
Creontiades pacificus	Brown mirid	3	Y	Y		Y	Y	
Whiteflys								
Trialeurodes vaporariorum	Greenhouse whitefly	3					Y	
Bemisia tabaci	Silverleaf whitefly	3					Y	
Bemisia tabaci	Eastern Aust. whitefly	3					Y	
True bugs								
Austroasca viridigrisea	Vegetable leafhopper	3					Y	
Amrasca terraereginae	Cotton leafhopper	3					Y	
Nezara viridula	Green vegetable bug	3		Y				
Piezodorus hybneri	Red-banded shield bug	3					Y	
Beetles								
Agrypnus variabilis	True wireworm	3						Y
Pterohelaeus spp.	False wireworms	3						Y
Nisotra spp.	Redheaded flea beetle	3					Y	
Chaetocnema spp.	Brown flea beetle	3					Y	
Monolepta australis	Redshouldered beetle	3	Y	Y			Y	
Carpophilus spp.	Flower beetles	3			Y			

Table 1.1: Australian cotton crop entomology.
The common pests of cotton in Australia, along with severity of damage that they cause on a scale of 1 to 3, with 1 representing serious annual damage and 3 representing only occasional damage. The tissue specificity of the damage cause by pests is indicated with a "Y" (Yes) in the column corresponding to tissue type. ("Term." corresponds to terminal region and "Seed." corresponds to seedling). Adapted from Fitt, (1994) and Pyke and Brown (1996).

generation time of both species is approximately 42 days, with a 17 day larval stage followed by a 16 day pupal stage (Pyke and Brown, 1996). Eggs of both species are laid on the top third of the plant, with a four-day incubation period before hatching. After hatching, the larvae (caterpillars) move down the plant and feed preferentially on the reproductive tissues and bolls (Matthews and Tunstall, 1994). Individual insects often eat through the outer wall of the cotton boll and consume the fibres, causing direct physical damage to the fibres as well as indirect damage by exposing the inside of the boll to fungal infection. Seedlings and the terminal region (young upper main stem with expanding leaves and squares) of maturing plants can also be attacked. These pests are a significant problem because of their high abundance in cotton regions and their preferential feeding on the commercially important part of the cotton crop (Fitt, 1994, Pyke and Brown, 1996).

H. punctigera is an obligate migrant that moves large distances every generation and is unique to Australia (Pyke and Brown, 1996). Populations breed in outback areas after winter rains, and in the spring, insects migrate to the coastal cropping areas where cultivated crops and weeds are initially infested, with subsequent generations infesting cotton. The timing of the *H. punctigera* population migration makes it an early season cotton pest, with few individuals present during the later part of the cotton-growing season. As well as this annually migrating population, there is also a small population that survives the cooler months of the year, between growing seasons, by entering a period of hibernation below the soil in pupal burrows. While some individuals emerge from this hibernation period, known as overwintering, the seasonally migrating population is of more importance (Sequeira, 2001). This is due to the larger size of the migrating population, which contains individuals that are more mature when they reach the cotton areas in early spring, compared to the overwintering population which emerges later (Pyke and Brown, 1996).

Large populations of *H. armigera* are found on cotton crops two to four weeks later than *H. punctigera*. However, within one to two generations, *H. armigera* is the more significant pest, with large population numbers persisting until late in the growing season (Fitt, 1994). Unlike *H. punctigera, H. armigera* does not show any major population-wide migration, with individuals usually remaining within five to ten kilometres of their hatching location (Farrow and Daly, 1987).

Spider mites (*Tetranychus spp.*) are also important pests in some cotton growing regions, with three species identified on cotton (Wilson, 1993). The two-spotted mite, (*Tetranychus urticae*) is the most abundant, while the bean spider mite (*Tetranychus ludeni*)

and the strawberry spider mite (*Tetranychus lambi*) are only rarely found. These species inhabit the underside of leaves and feed primarily upon mature leaves. When population numbers are very high, plant defoliation is possible and can lead to reduced fibre yield and quality. However, mites do not usually cause any serious damage because of predation by thrips and both *Helicoverpa* species (Pyke and Brown, 1996).

As well as the three primary pests of Australian cotton, there are also secondary pests including aphids, mirids, thrips and whiteflies, which can cause some damage in some regions in some seasons. Aphids, like spider mites, have a feeding preference for the leaves of the plant and only become a problem when their numbers are very high (Pyke and Brown, 1996). As well as damage from direct feeding, aphids can cause a reduction in fibre quality as a result of their honeydew excretion, which can cause stickiness and discolouration of the fibre. There are several species of mirid that are found in Australian cotton fields that can cause minor damage to the terminal region of seedlings, small squares and bolls as well as young leaves. Mirids feed by piercing plant tissues with their sharp mouthparts, releasing a chemical (pectinase) that destroys nearby tissues (Pyke and Brown, 1996). Thrips are a common early season secondary pest infesting the underside of cotyledons, young leaves and also the terminal region of the plant, but their damage is usually minor and rarely affects yield (Fitt, 1994). The whitefly can also be a significant pest of Australian cotton, causing crop damage by leaf feeding or by secretion of honeydew. However, levels of damage are rarely significant as the whitefly requires warm winter months and continuous availability of suitable host plants (Pyke and Brown, 1996).

There are numerous other pests that are capable of causing cotton crop damage but are not considered to be important pests because they only infrequently cause minor damage in isolated regions or to the borders of the cotton fields. These minor pests include the cluster caterpillar, the cotton looper and leafhoppers, which can cause plant defoliation by consuming leaves; armyworms that attack seedling cotton; the green vegetable bug and rough bollworms, which feed on squares and young bolls and the cotton tipworm, which feeds on young plant terminals (Fitt, 1994) (table 1.1).

A major consideration in determining the importance of a pest is what part of the plant is attacked (table 1.1). As discussed in section 1.5, minor damage to leaves and reproductive tissue does not cause significant reductions in fibre yield or quality because of the plants ability to tolerate and compensate for damage (Sadras, 1995, Bednarz and Roberts, 2001). However, major damage to reproductive tissues will result in reduced fibre yield and/or

quality; severe seedling damage can result in plant death and damage to the terminal region of the plant can hinder plant growth (Sadras, 1995).

Australian cotton beneficials

The classification of insects and arachnids as pest or beneficial inhabitants of the cotton crop is complex, since many individual species can have both negative and positive impacts on the crop at varying phases of plant development (Pyke and Brown, 1996). The major beneficial insects are those that feed on *Helicoverpa* pests at the egg, larval or adult (moth) stage and include: assassin bugs, big-eyed bugs, brown lacewings, brown smudge bugs, common brown earwigs, damsel bugs, glossy shield bugs, green carab beetles, ladybirds, predatory shield bugs, red and blue beetles and spiders. Another important group of beneficials that help to control *Helicoverpa* populations are parasitic wasps, both native and introduced.

Although less well characterised, there are also beneficial insects that feed on pests other than *Helicoverpa*. Some of these have specific feeding preferences such as the apple dimpling and big-eyed bugs, which feed on mites or brown smudge bugs; hoverflies, which feed on aphids; pirate bugs, which feed on thrips, and assassin bugs which feed on mirids. There are also generalist feeders that often feed on other beneficials as well as pests. These include: ants, brown and green lacewings, damsel bugs, glossy shield bugs and spiders.

1.5: Integrated pest management

The traditional method of dealing with pests that cause crop damage has been the application of broad-acting pesticides, with the aim of protecting the crop by killing as much of the pest population as possible. However, this approach has been expensive, both in economic and environmental terms (Fitt *et al.*, 2004). Pesticides have allowed high cotton yields but often kill beneficials and promote resistance in pest populations. This has resulted in the application of greater volumes of pesticides or the use of newer, more expensive pesticides. Large-scale aerial pesticide application has resulted in drift to nearby non-target crops, chemical residues in soil and water systems, and undesirable effects of the pesticide on neutral and beneficial insects as well as other wildlife.

An alternative to this traditional method has been to undertake a more subtle approach to dealing with the problems posed by pests. Known as integrated pest management (IPM), this approach has been defined as “the careful consideration of all available pest control techniques and subsequent integration of appropriate measures that discourage the

development of pest populations and keep pesticides and other interventions to levels that are economically justified and reduce or minimise risks to human health and the environment. IPM emphasises the growth of a healthy crop with the least possible disruption to agro-ecosystems and encourages natural pest control mechanisms" (Fitt *et al.*, 2004). Essentially this involves monitoring pest population dynamics and intervening with environmentally friendly, narrow-acting anti-pest treatments only when pest numbers reach a threshold level that could cause a commercially significant level of damage. The overall aim is to maintain profitability, yield and quality while reducing pesticide usage. One specific IPM practice is the cultivation of soil under the cotton crop between growing seasons to disturb the soil sufficiently to destroy overwintering pupae. The elimination of the overwintering population is beneficial because this population may have increased resistance to the endotoxin compared to non-overwintering individuals (Fitt *et al.*, 2004). Methods to conserve, augment or manipulate beneficial insect populations are also used as part of the IPM system (Fitt *et al.*, 2004).

Cotton plants produce more flower buds than can be sustained through to boll maturation. Under optimal conditions it is estimated that bolls are retained at only half of all potential fruiting sites (Bednarz and Roberts, 2001). When bolls are damaged, they are shed by the plant so that metabolic resources are not wasted on a boll which will not complete maturation (Fitt, 1994). Therefore, under moderate pest infestation levels, up to half of the plants bolls can be shed with no effect on fibre yield. In addition to producing excess fruiting sites, metabolic resources can be channelled away from damaged bolls to healthy bolls, allowing for the production of larger bolls with higher fibre yields (Sadras, 1995). Metabolic resources can also be reallocated into vegetative growth and the formation of extra fruiting structures (Kennedy *et al.*, 1986, Fitt, 1994). The capacity of the plant to compensate for damage has been recognised by IPM practices, meaning that a moderate amount of pest attack is tolerated and anti-pest treatments delayed or reduced.

1.6: Plant breeding

Traditional breeding has resulted in cotton cultivars that exhibit improved naturally occurring traits. Plant breeders have utilised desirable traits from different cotton species and cultivars by hybridising commercially grown cultivars with each other and also with wild cotton varieties. This directed breeding has allowed for desirable traits to be transferred between cotton varieties and has resulted in cultivars with improved characteristics (Wilkins *et al.*, 2000). Examples of traits introduced to commercial cotton varieties include resistance to key diseases (e.g. bacterial blight and Verticillium wilt) as well as morphological (e.g. okra

leaf shape and smooth leaf phenotypes) and biochemical (e.g. increased levels of tannins and terpenoids) defences against pests (Brook *et al.*, 1992, Wilson, 1994). The production of excess flower buds to compensate for damage has also been increased (Wilson *et al.*, 2003).

While directed breeding of cotton has generated enhanced cotton cultivars, there are several shortcomings to this method of cotton improvement. The process is expensive in terms of the time and resources required for plant crossing, growth and selection, with up to ten years required from initial hybridisation to commercialisation (John, 1997). After a desirable gene has been introduced into a species, several rounds of backcrossing are required to restore the commercial genetic background. Additionally, the need for genetic hybridisation limits the breeding to genetically compatible species that are capable of producing fertile progeny (John, 1997).

1.7: Plant biotechnology

An alternative to traditional breeding programs is the use of genetic engineering where genes that confer improved traits are introduced into the cotton genome. Several methods of transforming cotton are available including particle bombardment and *Agrobacterium*-mediated DNA transfer (reviewed by Wilkins *et al*., 2000). Once the new genetic material is introduced to individual cells, whole fertile plants are regenerated in tissue culture. Somatic embryogenesis during tissue culture is the most common technique and allows for the generation of whole transgenic plants after approximately 12 months. However, most elite cotton cultivars have low regeneration rates from somatic embryos, which means that cultivars with higher regeneration rates, usually those that are poorly performing as commercial varieties, must be used. This necessitates backcrossing, with up to several years required to recover an elite genotype. These techniques have been used to engineer several new cotton varieties, the vast majority of which have been engineered to be resistant to herbicides and pests (Willmitzer, 1999).

One of the main advantages of using genetic engineering rather than conventional breeding to generate new cotton varieties is the reduced time that is required: approximately three years compared with up to ten years for conventional breeding (John, 1997, Wilkins *et al.*, 2000). Transgenic cotton varieties have been readily accepted by the Australian cotton industry, with over 85% of the crop being planted to genetically modified varieties with insect and/or herbicide tolerance. This has allowed for a large reduction in herbicide and pesticide use, reduced weed and insect-associated crop losses and increased cotton fibre yield while reducing production costs (reviewed by Wilkins *et al*., 2000).

1.8: Pest-resistant cotton

The cotton crop is particularly susceptible to insect attack, due to its long period of fruiting, the ease at which cotton bolls are damaged and shed by the plant and the feeding preference of cotton pests for the commercially important part of the plant (Fitt, 1994, Pyke and Brown, 1996). The problem of cotton crop losses to pest attack is so great that up to 25% of world insecticides (Pannetier *et al.*, 1997) and 10% of world pesticides are used on cotton (Fitt, 1994).

Prior to 1996, the only method available for large-scale prevention of crop damage by pests was to apply pesticides. Large quantities of these chemicals were used to protect the crop, but pesticide application has several problems. Multiple applications are required every season, which is both expensive and time consuming. Pesticides are generally aerially applied, meaning that their application is affected by weather conditions. Furthermore, after application, factors such as rain and UV irradiation reduce the effectiveness of the pesticide (de Maagd *et al.*, 1999). Arial application also means that not all surfaces of the plant receive pesticide, and pests can survive by sheltering in or under plant tissues. Pesticides generally exhibit non-specific toxicity and can drift to non-target areas or be washed off of the farm by rain (de Maagd *et al.*, 1999). In addition, the continued application of broad-spectrum pesticides can lead to the developent of resistance in pest populations. In response to these economic and environmental concerns, pest-resistant transgenic cotton varieties have been developed.

1.8.1: *Bacillus thuringiensis*

Background

Of the numerous different pesticides that have been used on cotton crops, some of the most effective have been formulations manufactured from the *Bacillus thuringiensis* (*Bt*) soil bacterium (Berliner, 1915). Utilising the natural insecticidal properties of *B. thuringiensis* (variety *kurstaki*), insecticides have been produced by fermentation of the bacterium for use on many different crop plants since the 1930s (Shelton *et al.*, 2002). These have been effective against lepidopteran cotton pests, while not harming neutral or beneficial organisms (McClintock *et al.*, 1995). However, *Bt*-insecticides have several key drawbacks, most importantly they are short acting and non-penetrative, meaning that they are quickly degraded by environmental conditions and are unable to penetrate plant canopies.

The insecticidal activity of *B. thuringiensis* is derived from the activity of proteins that are included in crystals formed during bacterial sporulation. The genes encoding these toxic

proteins were are termed '*cry*' genes because of the 'crystal' proteins that they encode (Schnepf and Whiteley, 1981). When consumed by susceptible insects, the crystals of protein are solubilised and the Cry proteins are activated via cleavage by midgut proteases, under appropriate pH conditions (Gill *et al.*, 1992). The activated toxin can then bind to specific receptors on the surface of the insect's epithelial midgut cells. Once bound to the receptor, the toxin creates pores in the midgut epithelium causing fatal osmolysis (Gill *et al.*, 1992). This pore formation also disrupts the pH gradient in the midgut lumen and destroys cells. The gut becomes paralysed and the insect stops feeding, with subsequent death within several hours of toxin ingestion (Marrone and MacIntosh, 1993). The specificity of binding of Cry proteins to insect epithelial receptors is believed to mediate the ranges of toxicity that are observed for various Cry proteins (Hofmann *et al.*, 1988).

Different *cry* genes are toxic to different classes of insect, with more than 150 *cry* gene sequences classified from *B. thuringiensis* to date (*Bacillus thuringiensis* toxin nomenclature[3]) (Crickmore *et al.*, 1998). The *cry* genes are classified according to their evolutionary relationships and those characterised to date are termed *cry1* to *cry49* with further family and subfamily divisions. The proteins toxic for lepidopteran insects belong to the Cry1, Cry2 and Cry9 groups.

Development of the first generation of Bollgard®

The ineffective nature of sprayed insecticides, combined with extensive knowledge of *B. thuringiensis*, led to the development of *Bt*-transgenic cotton, known as Bollgard®, by Monsanto for the US market. Of the many *cry* genes present in the *B. thuringiensis* genome, *cry1Ac* was identified as encoding a protein with the most appropriate strength and range of toxicity for use as a cotton anti-pest transgene (Macintosh *et al.*, 1990). Several modifications to the gene sequence were required for high-level plant expression of the bacterial gene (reviewed by Perlak *et al.*, 1991). The resulting synthetic *cry1Ac*-like coding sequence was placed under the control of an enhanced constitutive *35S* cauliflower mosaic virus (*35S*) promoter. This produced high levels of expression of the toxin, such that it represents approximately 0.3% of total soluble protein throughout most plant tissues (Perlak *et al.*, 1990, Wilkins *et al.*, 2000).

Bt-transgenic cotton was released in the US in 1996 and in subsequent years throughout the rest of the world. Worldwide, Bollgard® was successful in providing good

[3] *Bacillus thuringiensis* Toxin Nomenclature (http://www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/index.html)

control of several lepidopteran pests including *H. armigera*, *H. punctigera,* tobacco budworm (*Heliothis virescens*), pink bollworms (*Pectinophora spp*.), rough bollworm (*Earias heugeli*), cotton tipworm (*Crocidosema plebejana*) and cotton looper (*Anomis flava*). Moderate control was provided for cotton bollworm (*Helicoverpa zea*), cabbage looper (*Trichoplusia ni*), saltmarsh caterpillar (*Estigmene acrea*) and cotton leaf perforator (*Bucculatrix thurberiella).* However, Bollgard® was not effective against several secondary pests including armyworms (*Spodoptera spp.),* soybean looper (*Pseudoplusia includens*) and cutworms (*Agrotis spp*.). As expected, the transgenic varieties were observed to have no effect on non-lepidopteran pests (reviewed by: de Maagd *et al*., 1999; Edge *et al*., 2001; Perlak *et al*., 2001; Sharma *et al*., 2004)

As with sprayed pesticides, the potential for insect populations to develop resistance exists with the deployment of transgenic insecticidal cotton. The evolution of resistance is dependant upon several factors, including the mode of action of the toxin, initial resistance frequency, generation time of the target insect, persistence of the toxin and lifecycle parameters of pest (Pimentel and Burgess, 1985). In the case of *Bt*-cotton, it is believed that the risk of pests developing resistance is low. This is based on the observation that the use of *Bt*-insecticidal sprays has rarely induced resistance, despite more than 50 years of use, with up to 70 applications per year in some areas (Marrone and MacIntosh, 1993). Additionally, it is believed that resistance alleles are rare, recessive and carry a fitness cost (Carriere *et al.*, 2001a, Carriere *et al.*, 2001b).

Before the commercial release of *Bt*-transgenic cotton varieties, a resistance management program was established. The long-held ideal of killing all individuals of a pest population was deemed unachievable because the population would eventually develop resistance under high selective pressure (Gould, 1994). Therefore, the concept behind the *Bt*-transgenic cotton resistance management program was to keep insect numbers below a level that could cause significant commercial damage. A program was implemented that required farmers to plant a portion of their crop area to conventional, non *Bt*-sprayed cotton, termed a refuge (Roush, 1998). This refuge allows the survival of some susceptible individuals, which can then breed with resistant individuals, so that the resistance alleles will be diluted in the pest population. Various refuge requirements were implemented in different countries depending upon several factors such as local pest characteristics, with the goal of preventing population-wide insect resistance.

_Bt_-transgenic cotton (Ingard®) in Australia

_Bt_-transgenic varieties, known as Ingard®, were developed for Australian conditions by the CSIRO and commercially released in 1997. These varieties provided very good protection from *H. punctigera* but were less effective against *H. armigera*. It was determined that the Cry1Ac protein was less toxic to *H. armigera*, and some insects were able to survive by feeding on the pollen and floral structures of the plant, where endotoxin expression is relatively low (Greenplate, 1999). Another problem with the Ingard® varieties was decreased efficacy as the season progressed, due to reduced endotoxin expression with increasing plant age (Fitt, 1994, Fitt *et al.*, 1998, Adamczyk *et al.*, 2001a, Adamczyk and Sumerford, 2001). Ingard® was generally effective against other lepidopteran pests (table 1.1), except for armyworms (*Spodoptera spp.*) and cutworms (*Agrotis spp.*) (Adamczyk *et al.*, 2001a).

Despite the problems with Ingard® cotton, population sizes of the major Australian pests were kept low and the total number of pesticide applications was reduced by ~50% (Fitt, 2004a). During the 2001/02 cotton-growing season, average total insecticidal applications targeted against *Helicoverpa* were reduced from 8.2 on conventional cotton to 1.6 on Ingard varieties (CRDC field trials, 2003[4]). However, a slight increase in the number of pesticide applications was required to control secondary pests (such as thrips, mirids, mites, whitefly and green vegetable bugs) on Ingard® cotton compared to conventional cotton, presumably because the pesticide applications on conventional cotton targeted against *Helicoverpa* were also controlling secondary pests.

The development of pest-population resistance to the endotoxin was a concern associated with the release of Ingard® varieties in Australia because of the high-level of _Bt_-endotoxin expression. As previously described, a large population of *H. punctigera* individuals migrate into cotton growing regions from central Australia every year. This migrating population has not been previously exposed to the _Bt_-endotoxin and carries very little resistance. Therefore, the migrating population acts to dilute the resistance alleles in the pest population, reducing the likelihood of population-wide resistance (Pyke and Brown, 1996). However, endotoxin resistance is much more likely for *H. armigera* populations since individuals remain within five to ten kilometres of their hatching location. It has also been observed that *H. armigera* individuals can survive on Ingard® crops by feeding on the pollen of the plant, a tissue that has relatively low endotoxin expression (Greenplate, 1999). This may increase the likelihood of populations developing resistance, since the population receives a sub-lethal dose of endotoxin, allowing individuals with higher levels of resistance

[4] Cotton Research and Development Corporation (http://www.crdc.com.au)

to survive (Frutos *et al.*, 1999). Therefore IPM approaches such as pupae busting are required to prevent individuals with resistance alleles surviving from one season to the next (Roush, 1997).

Within Australia, a high-dose/refuge resistance-management strategy was undertaken. This required *Bt*-transgenic plants to express a high level of endotoxin, capable of killing all susceptible insects, while a refuge of non *Bt*-transgenic plants provides a population of susceptible individuals to dilute the resistance alleles in the population (Fitt *et al.*, 2004). The refuge sizes were set based on computer modelling studies that assume resistance is recessive and conferred by alleles of a single gene. The model assumes 100% fatality of homozygous susceptible individuals and at least 95% fatality for heterozygotes that feed on the *Bt*-cotton. In numerous studies of *Bt*-resistance, the genes responsible for resistance are inherited in a recessive manner (Roush and Shelton, 1997). However, feeding studies with some pest species have shown that *Bt*-transgenic cotton does not meet the 95% heterozygote fatality levels that are required (Gould, 1998). If the assumptions of the modelling are shown to be inaccurate by experimental methods, then the recommended refuge sizes may not be sufficiently large to dilute the resistance gene in the insect population. Resistance to *Bt*-sprays has been observed under extreme application conditions but population-wide resistance has not been reported under normal field conditions (reviewed by Tabashnik *et al.*, 2004).

Development of Bollgard II®

A second generation of *Bt*-transgenic cotton, known as Bollgard II®, was developed to overcome the problems of Bollgard®/Ingard®. As well as the *cry1Ac* gene that was present in the first generation of *Bt*-cotton, the Bollgard II® varieties contain the *cry2Ab2* gene (reviewed by Perlak *et al.*, 2001). The protein encoded by *cry2Ab2* targets different receptors in the insect midgut than the Cry1Ac protein, meaning that insects resistant to one of the endotoxins are not resistant to the other (Perlak *et al.*, 2001). Population-wide resistance to the two endotoxins is therefore unlikely, as for individuals to survive they must possess resistance to both endotoxins, which induce toxicity through distinct modes of action (Alcantara *et al.*, 2004).

Worldwide, Bollgard II® was more successful than the first generation of *Bt*-transgenic cotton due to an increased level and range of toxicity (reviewed by Sharma *et al.*, 2004). Improved promoter technology allowed the Cry2Ab2 protein in Bollgard II® to be expressed at eight times the level of the Cry1Ac protein in Bollgard (Adamczyk *et al.*, 2001b). The additive effect of the two endotoxins with distinct modes of toxicity increased

toxicity to species that were not adequately controlled by Bollgard®, such as *Helicoverpa spp., Pectinophora spp., Spodoptera spp.* and *Agrotis spp.* (Perlak *et al.*, 2001). However, as was the case for Bollgard®, Bollgard II® varieties also have declining efficacy with plant age, allowing some susceptible individuals to survive until late in the growing season (Adamczyk *et al.*, 2001b).

Commercial trials by Monsanto (Monsanto field trials, 2002)[5] have demonstrated large reductions in insecticide use for Bollgard II® compared with Bollgard®/Ingard® and conventional cotton (figure 1.2). Total applications required to control pests were decreased from an average of ten per season for conventional cotton to four per season for the first generation of *Bt*-cotton (Bollgard® and Ingard®), and two per season for the second generation of *Bt*-cotton (Bollgard II®). Developing countries such as China have seen similar reductions, with 60% to 80% fewer insecticidal applications required to control pests for Bollgard II® varieties relative to conventional varieties (Fitt, 2004b, Wu and Guo, 2004).

Within Australia, Ingard® varieties were grown commercially from the 1996/97 season until the 2003/04 season, with the new Bollgard II® varieties completely replacing the Ingard® varieties after 2003/04. In 2005/06, Bollgard II® accounted for approximately 80% of the cotton crop area (CRDC, 2006[6]). The high level of toxicity of Bollgard II® towards *Helicoverpa* has resulted in a ~80% reduction in the number of pesticide applications, affording large financial and environmental benefits (Cotton Australia, 2004[7]). As Bollgard II® provides better control of lepidopteran insects, fewer broad-spectrum pesticide applications are required, which has led to increased numbers of several secondary pests (e.g. mirids, green stink bugs, green vegetable bugs). However, other secondary pests (e.g. mites and aphids) have been well controlled on *Bt*-cotton by the increased activity of beneficial insects (Fitt *et al.*, 2004).

Other genes encoding anti-pest molecules

While *Bt*-cotton was being developed, research was also being conducted to identify other genes that encode proteins with insecticidal properties and could be used as anti-pest transgenes. The most promising are discussed below while information on other groups of anti-pest genes including lectins, enzymes (cholesterol oxidase, polyphenol oxidases and peroxidases, chlorogenic acid, lipoxygenase, bacterial isopentenyl transferase and chitinase),

[5] Monsanto Company (http://www.monsanto.com)
[6] Cotton Research and Development Corporation (http://www.crdc.com.au)
[7] Cotton Australia (http://cottonaustralia.com.au)

NOTE: This figure is included on page 17 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.2: Pesticide applications on cotton varieties in 2002.
The number of pesticide applications used to control lepidopteran and nonlepidopteran pests on conventional, Ingard® and Bollgard II® cotton (Monsanto field trials, 2002).

secondary plant metabolites (alkaloids, steroids, foliar phenolic esters terpenoids, saponins and flavonoids) and toxins from predators (scorpion and spider venom genes) can be found in reviews elsewhere (Sharma *et al.*, 2000, Carlini and Grossi-de-Sa, 2002, Babu *et al.*, 2003, Sharma *et al.*, 2004, O'Callaghan *et al.*, 2005).

1.8.2: Vegetative insecticidal proteins

Because of the success of the Cry proteins, supernatant fluids from the *B. thuringiensis* bacterium were screened for other proteins with insecticidal properties. From these screens, a group of proteins, termed vegetative insecticidal proteins (VIP), were isolated (Estruch et al., 1997). The toxicity of the VIPs is approximately the same as the Cry proteins, with a similar mode of action. As with the Cry proteins, VIPs undergo a proteolytic cleavage in the insect midgut (Yu et al. 1997) and once activated, bind to specific receptors inside the insect midgut epithelium of susceptible insects. Upon binding, the protein inserts into the membrane resulting in progressive degeneration of the epithelial layer and the formation of ion-specific pores, causing disrupted digestion and subsequent death (Lee et al., 2003). Transgenic cotton varieties expressing VIPs have been developed and are now at the field trial stage. Within Australia, a cotton variety expressing VIP3A is being grown in field trials and may be commercialised in coming years (OGTR, 2006[8]).

1.8.3: Enzyme inhibitors

Protease inhibitors

Insects utilise a range of different proteases to digest dietary proteins into amino acids, which are required for insect metabolism. Several classes of protease inhibitor have been used as anti-pest transgenes in various plant species (Thomas *et al.*, 1995, Sharma *et al.*, 2004, O'Callaghan *et al.*, 2005). Most protease inhibitors are unsuitable for use as anti-pest transgenes because they have a wide range of toxicity and therefore affect many insect species. Since lepidopteran insects rely mainly on serine proteases for protein digestion (Boulter, 1993) several plant species, including cotton (OGTR, 2003[9]), have been engineered with serine protease inhibitors. The typical effects of protease inhibitors are mildly increased larval mortality and reduced larval growth rates. These non-lethal modes-of-action have generally been shown to be ineffective in field trials because pests are able to tolerate the protease inhibitor by increasing levels of other classes of protease or by simply over-producing the inhibited protease (Sharma *et al.*, 2000, Sharma *et al.*, 2004). The problems associated with protease inhibitors could potentially be overcome by expressing multiple

[8] Office of the Gene Technology Regulator (http://www.ogtr.gov.au/ir/dir034.htm)
[9] Office of the Gene Technology Regulator (http://www.ogtr.gov.au/ir/dir048.htm)

protease inhibitors of different types or by improving the affinity of introduced protease inhibitors for the target insect proteases (Michaud, 1997).

Alpha-amylase inhibitors

Alpha-amylases catalyse the hydrolysis of α-D-1-4-glucan linkages in starch, glycogen and other carbohydrates required for insect metabolism. Therefore, alpha-amylase inhibitors have been characterised and used to engineer transgenic plants, such as tobacco, for improved insect resistance (Sharma *et al.*, 2000). Research by Giri and Kachole (1998) demonstrated that alpha-amylase inhibitors from pigeon pea inhibited 22% of amylase activity in *H. armigera*. However, as is the case for protease inhibitors, most insect species contain a complex mix of amylases in the midgut lumen, and can react to inhibitors by overproducing other amylases. In addition, some insects also have a serine protease that can cleave alpha-amylase inhibitors (Ishimoto and Chrispeels, 1996). There are also difficulties identifying inhibitors that have a broad enough range of toxicity to affect multiple insect species, whilst not inhibiting the plants own alpha-amylases or those of non target species (Franco *et al.*, 2002).

1.8.4: Cyclotides

A family of unusual proteins with cyclised head-to-tail peptide backbones have been characterised in diverse organisms from bacteria to mammals with wide-ranging biological activities (Felizmenio-Quimio *et al.*, 2001). Recently, members of this family with anti-pest properties have been identified in several plant species (Hernandez *et al.*, 2000, Jennings *et al.*, 2001, Jennings *et al.*, 2005). These proteins, known as cyclotides, are characterised by a circular backbone structure and a knotted arrangement of cross-linking disulphide bonds, which gives the proteins a very stable structure that is impervious to enzymatic breakdown (Chen *et al.*, 2005). One of these circular proteins is toxic to *H. punctigera* and causes retarded insect growth and development (Jennings *et al.*, 2001). The mechanism of activity is unknown but the toxicity of the circular proteins towards lepidoptera is greater than that of either protease or alpha-amylase inhibitors, which indicates potential for their future uses as anti-pest transgenes.

Summary

In Australia, the *cry* genes of *Bacillus thuringiensis* are the only transgenes that have been utilised to generate commercialised transgenic cotton. Varieties engineered with the *vip3a* gene are in advanced trials and may be released as commercial varieties in the future. Other potential transgenes such as those previously described have been investigated and

characterised but do not seem as promising as either *cry* or *vip* genes because they require much higher expression levels, cause chronic rather than acute effects and have less desirable ranges of toxicity (Estruch *et al.*, 1997, Bent and Yu, 1999, Sharma *et al.*, 2004). Despite the disadvantages associated with non-*Bt* transgenes, several studies have shown they can act synergistically with *Bt*-transgenes to increase plant resistance against insect attack. By combining anti-pest transgenes that act through different biochemical pathways, the evolution of insect resistance is much less likely (Boulter, 1993, Roush, 1997, Roush and Shelton, 1997, Zhao *et al.*, 2003, Sharma *et al.*, 2004, Mehlo *et al.*, 2005).

1.9: Transgene expression

The majority of commercially-grown transgenic crops are engineered with constitutive promoters driving transgene expression (Sunilkumar *et al.*, 2002a). The most common of these constitutive promoters are of viral origin, with the *35S* promoter being the most widely used. A number of promoters from housekeeping genes such as actin and ubiquitin from various plant species have also used to drive constitutive transgene expression (Schunmann *et al.*, 2003).

All current commercial and field trial- stage transgenic cotton varieties that express insect resistance genes utilise the constitutive *35S* promoter for transgene expression. While the *Bt*-transgenic cotton varieties have been very successful, this constitutive transgene expression may be unnecessary, as the major Australian cotton pests preferentially attack the flower and boll tissues of the plant causing minor, if any, damage to other plant structures. In addition, constant high-level expression of transgenes may have several detrimental effects, such as placing strong selective pressure on pest populations to develop resistance, non-target effects of the transgene on other organisms, a yield penalty to the plant, and the presence of transgenic protein in secondary commercial products. These problems could be minimised by using a promoter that drives tissue-specific transgene expression in only the boll wall.

One of the major concerns about the effectiveness of transgenic insecticidal crops is the potential for insect populations to develop resistance to the transgene. The lepidopteran pests of cotton feed predominantly on the flower and boll tissues with some minor feeding on other non-commercially important tissues such as leaves. Therefore, the exposure of pests to toxin in these tissues is not required and can only act to increase selective pressure. In addition, the expression patterns of the *Bt*-transgenes are less than ideal. Expression is low in pollen and overall plant expression decreases with plant age (Adamczyk *et al.*, 2001b). This expression profile leads to insects receiving only a moderate dose of endotoxin that is not

sufficient to induce lethality, conditions that are believed to encourage population-wide resistance (Gould, 1998, Frutos *et al.*, 1999). Restricted expression surrounding the commercially important tissues of the plant would reduce this selective pressure and act to minimise the risk of insects becoming resistant to transgenic crops.

Another concern with constitutive expression of transgenes is the increased exposure of non-target insects and the environment to the endotoxins. No studies accurately representing field conditions have demonstrated that Cry1Ac or Cry2Ab2 proteins have negative effects on non-target insects that directly feed on the plant, despite several preliminary reports of negative effects (Gatehouse, 2002). However, unexpected non-target effects have been observed for other Cry proteins such as the effect of Cry1Ab on the Egyptian cotton leafworm (*Spodoptera littoralis*) (Vojtech *et al.*, 2005). In addition, several studies have shown that beneficial predators and parasitic insects can be negatively affected by *Bt*-crops. For example, on *Bt*-corn, the parasitic wasp (*Tetrastichus howardi*) prefers to lay eggs in generalist herbivore *Chilo partellus* individuals that inhabit non *Bt*-corn rather than those that inhabit *Bt*-corn. *T. howardi* females derived from *C. partellus* fed on *Bt*-corn also have lower fecundity and shorter lifespans, thereby reducing their efficacy in controlling pest populations (Prutz *et al.*, 2004). Restricted transgene expression would limit exposure of neutral and beneficial insects to potentially harmful *Bt*-toxin.

In addition to effects on non-target insects, recent studies have assessed the level of Cry protein that is incorporated into the soil under *Bt*-crops. Cry protein can enter the soil when plant tissue breaks down within the soil or through the release of exudates from roots (Saxena and Stotzky, 2000). Estimates of soil concentrations of various Cry proteins from several crop species vary widely from 1.58 µg of Cry1Ac/kg of soil to 2410 µg of Cry3Bb1/kg of soil (reviewed by Clark *et al.*, 2005). Soil persistence studies of Cry proteins have also shown great variation in persistence of bioactivity ranging from a few days to six months, suggesting that long-term exposure to Cry proteins through soil residues is possible (reviewed by Clark *et al.*, 2005). *Bt*-proteins may have negative effects on earthworms, with several studies demonstrating weight loss in response to *Bt*-protein consumption (Zwahlen *et al.*, 2003). In addition, soil and rhizospheric microbial communities may be altered with exposure to plant-derived Cry proteins (reviewed by Castaldini *et al.*, 2005). Localised transgene expression would decrease Cry protein soil concentration by eliminating expression in most tissues of the plant and therefore reduce the exposure of soil dwelling organisms to potentially harmful endotoxins.

An additional advantage of restricted transgene expression may be a reduced metabolic cost to the plant. The proteins encoded by the *Bt*-transgenes in current transgenic cotton varieties account for approximately 1.5% of total soluble protein (Perlak *et al.*, 2001) and alternative potential anti-pest transgenes would require expression at even higher levels to overcome their relatively low levels of toxicity. High level expression of transgenes has been theorised to cause a yield penalty to the plant, due to the high metabolic expenditure on the production of the protein (Schuler *et al.*, 1998, Hilder and Boulter, 1999). Field studies to show this yield penalty have not been performed, perhaps due to the fact that the theoretical effect is mild and difficult to discern from the effects of genome disruption caused by transgene insertion. If transgene expression was localised to the tissues where it is required, then the theoretical metabolic cost would be reduced. This would allow the plant to allocate more resources to growth and fibre development, which may in turn result in increased plant health and fibre yields.

As well as potentially reducing the effects on non-target insects and increasing crop yields, secondary products would also be improved by restricted transgene expression. Secondary commercial products from the cotton crop are derived from the processed seed, and include oils used for human consumption and stock feed. Although processing eliminates up to 97% of the transgenic proteins from the seed oil (Sims and Berberich, 1996), transference of any amount of transgenic protein to human food is undesirable.

Research and development into the transgenic Ingard® and Bollgard II® cottons was conducted at a time when constitutive promoters were the best available tools for cotton transgene expression. However, in recent years there has been increased interest in other, more sophisticated methods of transgene expression using tissue-specific, time-specific, wound-inducible, chemical-inducible and novel synthetic promoters (reviewed by Schuler *et al.*, 1998; Lessard *et al.*, 2002; Gurr and Rushton, 2005). A *Bt*-corn variety known as KnockOut® was developed with transgene expression restricted to green tissues and pollen as driven by the phosphoenolpyruvate carboxylase (PEPC) (Hudspeth and Grula, 1989) and calcium-dependant protein kinase maize-derived promoters (Estruch *et al.*, 1994). The major advantage of this transgenic variety is the small amount of endotoxin expression in the commercially important tissue, the corn kernel (Koziel *et al.*, 1993, Sharma *et al.*, 2000). While the commercialised variety was successful in preventing insect attack, the high level of transgenic protein in the pollen may have been harmful to the monarch butterfly (*Danaus plexippus*) and the variety was withdrawn from the market. In addition to corn, a *Bt*-rice

variety has been engineered with the PEPC promoter driving green tissue-specific expression but has not yet been commercialised (Datta *et al.*, 1998, Husnain *et al.*, 2002).

1.10: Summary and Project aims

The theoretical advantages of restricted transgene expression mean that plants expressing transgenes such as *Bt*-endotoxins in a tissue-specific manner have the potential to be significantly better than the current constitutively expressing crops. Restricted expression could lead to increased yields, decreased environmental impact, decreased selective pressure on insects to develop resistance, and improved secondary products. Therefore, the broad aim of this study is to identify promoters that would be suitable for driving tissue-specific transgene expression in future transgenic varieties. Theoretically, the boll wall is an ideal tissue in which to localise transgene expression. The boll forms a physical barrier that protects the cotton fibres, and localised anti-pest transgene expression in this tissue would protect the fibres from insect attack. Since the major Australian cotton pests preferentially attack the boll, expression of anti-pest molecules in this tissue alone is likely to provide adequate protection of the fibres. In this project, a differential screen was utilised to identify mRNAs that are specifically transcribed in the boll wall. The corresponding promoters were identified, isolated and characterised. The expression patterns that these promoters drive were then analysed by performing transient transformations of various cotton tissues.

Chapter 2: Materials and Methods

2.1: Materials

2.1.1: Plant material

All cotton material was isolated from plants grown from seed provided by Cotton Seed Distributors (Narrabri, NSW). Cotton species and varieties used were:

G. hirsutum cultivar Siokra 1-4

G. hirsutum cultivar Sicot 189

G. hirsutum cultivar Sicot V-2

G. raimondii cultivar

G. herbaceum variety *africanum*

2.1.2: Enzymes

Calf intestinal phosphatase (CIP)	Roche
DNase I (RNase free)	Roche
Klenow fragment of DNA polymerase I	Geneworks and MBI Fermentas
M-MLV reverse transcriptase	Promega
Pfu Ultra Hotstart DNA polymerase	Stratagene
Proteinase K	Roche
Restriction endonucleases	New England Biolabs (NEB) and Promega
RNasin	Promega
RNase A (Ribonuclease A)	Sigma-Aldrich
Shrimp Alkaline Phosphatase (SAP)	Amersham-Pharmacia
Spermidine	Sigma-Aldrich
T4 DNA ligase	NEB and Promega
Taq DNA polymerase	NEB and Geneworks

2.1.3: Radioactive isotopes

[α-^{32}P] dATP(3000 Ci/mMol)	Perkin-Elmer
[α-^{32}P] dCTP(~400 Ci/mMol)	Perkin-Elmer

2.1.4: Molecular weight markers

λ DNA restricted with *Hin*dIII	MBI Fermentas
2-Log DNA ladder	NEB
RNA Marker, 0.28-6.58 kb	Promega
1 kb Plus DNA ladder	Invitrogen

2.1.5: Antibiotics and indicators

Ampicillin	Roche
Chloramphenicol	Sigma-Aldrich
Kanamycin	Sigma-Aldrich
IPTG (isopropyl-beta-D-thiogalactopyranoside)	Progen
X-GAL (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside)	Progen
X-GLUC (5-bromo-4-chloro-3-indolyl-beta-D-glucuronide cyclohexylamine salt)	Progen

2.1.6: Kits

GenElute Plasmid Miniprep Kit	Sigma-Aldrich
mRNA Purification Kit	Amersham-Pharmacia
Packagene λDNA Packaging System	Promega
pGEM-T Easy	Promega
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
QIAquick Plasmid Midiprep Kit	Qiagen
Superscript Choice System for cDNA Synthesis	Invitrogen
Universal Genome Walker® Kit	Clontech

2.1.7: Stains and dyes

Bromophenol blue	Sigma-Aldrich
Ethidium bromide	Sigma-Aldrich

2.1.8: Chemicals reagents and miscellaneous

General laboratory chemicals were of analytical research grade and were purchased from a variety of manufacturers. Specialist reagents used in this study are as follows:

Antifoam A emulsion	Sigma-Aldrich
β-mercaptoethanol	Sigma-Aldrich
Bio-Gel P60 (100-200 mesh)	Bio-Rad Laboratories
Bio-Gel P60 (50-100 mesh)	Bio-Rad Laboratories
Blotting and filter papers	Whatman
DEPC (diethylpyrocarbonate)	Sigma-Aldrich

DMSO (dimethyl sulfoxide)	BDH Chemicals
Ficoll	Sigma-Aldrich
Fuji RX Medical X-ray film	Fuji
Gold particles (1.0 micron)	Bio-Rad Laboratories
Hybond-N^+ nylon membrane	Amersham-Pharmacia
Murashige and Skoog basal media	Sigma-Aldrich
Polyvinylpyrollidone (PVP)	Sigma-Aldrich
Salmon sperm DNA	Sigma-Aldrich
Sodium borate decahydrate	Sigma-Aldrich
Swinney 13 mm plastic filter holder	Pall Life Sciences
Triton X-100	Sigma-Aldrich
Miracloth	Calbiochem

2.1.9: Plasmid and bacteriophage vectors

pGEM-T Easy	Promega
pBluescript SK(-)	Stratagene
pMDC43GUS	(Curtis and Grossniklaus, 2003)
λZAPII *Eco*RI pre-digested and CIP treated	Stratagene
ExAssist helper phage	Stratagene
pJK.KiwiGUS.ocs.Kmf(-)	Lab Stock (Kirschman and Cramer, 1988)
pMDC162GUS	(Curtis and Grossniklaus, 2003)

2.1.10: *E. coli* Bacterial strains

DH-5αF'	Laboratory stock
One Shot competent cells	Invitrogen
SOLR	Stratagene
XL1-Blue MRF'	Stratagene

2.1.11: Oligonucleotides

Oligonucleotides were purchased from Geneworks (except where indicated) and were of standard PCR/sequencing grade. Primers were designed using NetPrimer (http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html) and are shown below as 5’ to 3’ sequences.

Primer	Sequence (5' to 3')
General primers	
Mixed random decamers	Geneworks
Oligo $d(T)_{12-18}$	Amersham-Pharmacia
Oligo $d(A)_{40-60}$	Amersham-Pharmacia
SP6	GATTTAGGTGACACTTAG
T7	AATACGACTCACTATAG
M13F	GTAAAACGACGGCCAGT
M13R	GTTTTCCGAGTCACGAC
RSP	AACAGCTATGACCATG
Genome Walker® PCR primers	
PRP-GW1-GSP1	GAGTAGGCGGGACTACAGGTGAAACAGC
PRP-GW1-GSP2	GCATGGTAATGAGGAGGAGAAGCCTG
PRP-GW2-GSP1	GCAAAGTTTCCAAGGAGCACCAATTTC
CHS-GW1-GSP1	CAACACAATTAGGTGGGGTTGATGTGC
CHS-GW2-GSP1	CTTGGGAGAGGGTAGTTGAGATTCACG
CHS-GW2-GSP2	TCCACATTTCTGTTGGTTCTCACTTACCG
GBSS-GW1-GSP1	CCTTTGCGTTGGTGGTCCTCATCTGC
GBSS-GW1-GSP2	CTATCGACCTTGTTCAAAGACCGCAAACC
GBSS-GW1-GSP3	TGATCGAAAGCTCAAGAACTCGAGATGG
Promoter amplification primers	
PRP-Forward	CACCCACAACCTTTCTAATCGTTAATGCC
PRP-Reverse	GGTAATGTTTAGAGACAAAGAAAGAAGATGG
PRP-Forward-Kpn	CGCGGTACCCCCACAACCTTTCTAATCGTTAATGCC
PRP-Reverse-Kpn	CGGGGTACCGGTAATGTTTAGAGACAAAGAAAGA
CHS-Forward-Kpn	CGGGGTACCTAAAATATATTATGCTATAATATTAATTTATAACTTTC
CHS-Reverse1-Xba	CTAGTCTAGATTTCGGATGTACAAAAACAGCCAACG
GBSS-Forward-Kpn	CGGGGTACCATTAAATCGATTAATTCAACCATTG
GBSS-Reverse1-Xba	CTAGTCTAGATGATCGAAAGCTCAAGAACTCGAGATGG
35S-Forward-Kpn	CGGGGTACCGCAGGTCAACATGGTGGAGCAC
35S-Reverse-Kpn	CTCGGTACCCGGGGATCCTCTAGAGTCGAGG
Sequencing primers	
Proline rich protein	
GW5-1-Reverse	GGGCAAAATTCCTTCTGTCG
5GW2-ReForward	GTATGTAATCAATCATTTCG
5GW2-ReReverse	GACCCTAAAAGCCTACACC
5ReForward	CGGTGGCACTAAGAATTGG
5ReReverse	CCAGGTGGCAGTGCAGGGCG
5Forward1	CCCACAACCTTTCTAATCGTTAATGCC
PRP-Forward1A	ACGAAATAATAGGATTCCCACC
PRP-Forward3	ATTTTGTGAATATGCTATGGTAGCC
PRP-Reverse4	GTTGTCAAGCTAATGTTACATCC
Chalcone synthase	
7Forward	CTCATGATGTACCAACAAGG
7Re-Forward	CGATTATAACAGCTGCGGCACC
7Re-Reverse	GGATTGCTCATCCTGGTGGTCC
Granule-bound starch synthase	
28Forward	CTGTTCGGTTCTTCCACTGC
28Reverse2	CCACAACCTTGATGGTCCC
28Re-Forward	GCCAGATTGTAGTCCTTGGC
28Re-Reverse	CGAATGAACTCTTGAACCG
28Re-Reverse2	CCTTGTACTGATCATAGCGAGG
28GW-6.4-Forward	CAAATTAACATCTTCTTACC
28GW-6.4-Reverse	GAACCCACCTAAGACTGAAGC

Sucrose synthase	
11Forward	GAGTATCTTGGTACCCTACC
11Reverse	GCAGGCTGTACAAAGGCACC
11Reverse2	GGAAGTATATCTCCATATCAGC
11Forward2	GGAAATTGCAGGAAGCAAGG
11Reverse2#2	GGAAATGCTTCAACCTCCGC
11Reverse3	GGAACTTCTCAAGGGTGCAAGG
11Forward4	GCTGCTGACATACTGGTCG
Anthocyanidin reductase	
24Re-Forward	GCAATGAATTCCTCATAAATGC
24Re-Reverse	CCTATGCTGCTGGGGACAATTGG
Dihydroflavonol 4-reductase	
27Re-Forward	CCTCATAACTGCACTTTCTCC
27Re-Reverse	CCAACAGGTTTCATCATAACAGG
27Re-Reverse2	GCAACATCAATAGTTCCAGC
Actin	
26Forward	CCAATCTATGAAGGATATGC
GUS gene	
GUS-Forward	AGTCTGGATCGCGAAAACTGTG
GUS-1AForward	TGGTGATGTGGAGTATTGCC
GUS-4Forward	GTCGGCGGCTTTTCTGCTGC
GUS-2Reverse	AAATATTCCCGTGCACCTTGC
GUS-3Reverse	GCTAGTGCCTTGTCCAGTTGC
pMDC162GUS-Reverse	GTTAAAACTGCCTGGCACAGC

2.1.12: Solutions, buffers and media

All solutions and buffers were prepared using Millipore-filtered water and (when appropriate) were sterilised by autoclaving. Non-autoclavable solutions were sterilised by filtration through a 0.2 μm filter. Solutions for RNA work were treated with 0.1% (v/v) DEPC prior to autoclaving. Solutions and all other buffers routinely used in this study were as follows:

GUS assays

GUS prefixing solution	100 mM $NaPO_4$ buffer, 0.001% (v/v) 40% formaldehyde, 0.001% β-mercaptoethanol, 0.01% Triton X-100
GUS pre-assay buffer	0.5 mM Potassium ferricyanide, 0.5 mM Potassium ferrocyanide, 50 mM $NaPO_4$ buffer (pH 7.0), 0.1% (v/v) Triton X-100
GUS assay buffer	87% (v/v) GUS pre-assay buffer, 0.3% (v/v) X-Gluc, 0.1% (v/v) β-mercaptoethanol, 1% Chloramphenicol, 10% (v/v) DMSO

RNA manipulation

Hot borate buffer	0.2 M sodium-borate decahydrate, 30 mM EGTA, 1% (w/v) SDS, 1% sodium-deoxycholate
MOPS buffer (10x)	200 mM MOPS, 50 mM anhydrous sodium acetate, 10 mM EDTA, adjusted to pH 7.0
RNA loading buffer	50% (v/v) formamide, 17.5% formaldehyde, 10% (v/v) 10x MOPS buffer, 2.5g/L ethidium bromide

General

Denaturing solution	0.4 M NaOH
Denaturing solution	1.5 M NaCl, 0.5 M NaOH
Denhardt's solution (1x)	0.02% (w/v) Ficoll, 0.02% (w/v) PVP, 0.02% (w/v) gelatine
DNA loading buffer (6x)	0.25% (w/v) bromophenol blue, 40% (w/v) sucrose.
Lysing solution (Grunstein's)	10 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA, 4 mg/ml lysozyme
Neutralising solution	1.5 M NaCl, 1 M Tris-HCl (pH 7.2), 1 mM EDTA
Phenol/chloroform	50% (v/v) phenol, 48% (v/v) chloroform, 2% (v/v) isoamyl alcohol, buffered with an equal volume of Tris-HCl (pH 8.0)
Premade-hybridisation buffer	5x SSPE, 5x Denhardt's solution, 0.5% (w/v) SDS.
Hybridisation buffer	90% (v/v) premade-hybridisation buffer, 1.5 mg sonicated salmon sperm (denatured by 90^{o}C for five minutes)
TAE (1x)	40 mM Tris-HCl, 20 mM Na-acetate, 2 mM EDTA (adjusted to pH 7.8 with glacial acetic acid)
TE (1x)	10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)
Stop buffer	10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0), 0.2% (w/v) SDS
SSC (1x)	0.15 M NaCl, 0.015 M sodium citrate (adjusted to pH 7.2 with NaOH)
Sodium phosphate buffer	57.7 mM $Na_2HPO_4.7H_2O$, 42.3 mM $NaH_2PO_4.H_2O$. (pH 7.0)
SM buffer	100 mM NaCl, 10 mM MgSO4, 50 mM Tris-HCl (pH 7.5), 2% (w/v) gelatine

SSPE (1x)	0.18 mM NaCl, 10 mM NaH2PO4, 1 mM EDTA (pH 7.2)
TAE (1x)	40 mM Tris-HCl, 20 mM Na-acetate, 2 mM EDTA, (adjusted to pH 7.8 with glacial acetic acid)
TE (1x)	10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)

Media

All media were prepared with Millipore water and were autoclaved for 15 minutes at 121°C. Non-autoclavable medium components were filter-sterilised and added to the autoclaved basal medium under aseptic conditions prior to use. Media purchased from commercial suppliers were prepared in accordance with the manufacturer's instructions.

Luria bertini broth	1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract (pH 7.2)
Luria bertini agar	Luria broth with the addition of 1% (w/v) bacteriological agar
MS basal medium	1x Murashige and Skoog basal salt macronutrient solution; 1x Murashige and Skoog basal salt micronutrient solution, pH 5.8
½ MS agar	½ Murashige and Skoog basal medium with 0.75% agar
NZY agar	1% (w/v) NZ amine (casein hydrolysate), 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 10 mM $MgSO_4$, 1% (w/v) bacteriological agar No. 1
NZY top agar/agarose	1% (w/v) NZ amine (casein hydrolysate), 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 10 mM $MgSO_4$, 0.7% (w/v) bacteriological agar No. 1 or agarose
SM buffer	100 mM NaCl, 10 mM $MgSO_4$, 50 mM Tris-HCl (pH 7.5), 2% (w/v) gelatine
YT broth (2x)	1.6% (w/v) tryptone, 1% (w/v) NaCl, 1% (w/v) yeast extract (adjusted to pH 7.0 with 1 M NaOH)

2.2: Methods

Standard molecular methods were performed as described in Sambrook *et al.* (1989). All kits were used in accordance with the manufacturer's instructions. All plants were grown in PC2-approved facilities in accordance with guidelines issued by the Office of the Australian Gene Technology Regulator.

2.2.1: Plant techniques

Plant growth conditions

Cotton plants were maintained in growth cabinets at temperatures of 30°C (day) and 25°C (night) with a day/night cycle of 16/8 hours. Plants were illuminated by mercury vapour growth lamps and were grown in premium commercial potting mix at a density of two to four per eight litre pot. Plants were fertilised periodically with commercial fertilisers.

Plant tissue harvesting

Cotton flowers were tagged on the day of anthesis. Bolls and associated tissues were harvested at various days postanthesis (DPA) and stored on ice before dissection, followed by freezing in liquid nitrogen and storage at -80°C. Bract tissue was removed by cutting the bracts away from the boll, at the base of the bract. Calyx tissue was isolated after cutting the base of the boll, where the calyx joins the boll. Boll wall tissues were harvested by cutting open the boll along the carpel wall lines, followed by removal of fibre and seed tissues. The white film tissue (a thin film that separates the fibres from the inside of the boll) was removed from younger bolls by peeling it away from the inside of the boll. Fibres were removed from the seeds under cold dH_2O using fine forceps and blotted on paper towel. Care was taken to avoid removal of the fuzz hairs or the seed coat. Petal and staminal column tissues were excised on the day of flowering. Root, stem and cotyledon tissues were harvested from seedlings of approximately seven centimetres in height.

2.2.2: General molecular techniques

Isolation of genomic DNA

Total genomic DNA was isolated from the young leaves of cotton by the method of Dellaporta *et al*., 1985.

Plasmid isolation

Plasmid DNA for all general applications was isolated from 10 ml of overnight culture by using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich) or by the alkaline lysis method

(Sambrook *et al.*, 1989). Large-scale plasmid preparations for use in microprojectile bombardment were performed with the QIAquick Plasmid Midiprep Kit (Qiagen).

RNA isolation

Total RNA was isolated from various cotton tissues using the method described by Wan and Wilkins (1994), except that all reagents were doubled to accommodate 1 g of tissue and a polytron was used in place of a glass tissue grinder. RNA was precipitated and stored in 0.1 volumes of 3 M sodium acetate and 2 volumes of 100% ethanol at -70^{o}C.

Determination of DNA and RNA concentration

The concentration of DNA and RNA in solution was determined by UV spectrophotometry at a wavelength of 260 ηm, or by gel electrophoresis of a sample and comparison with *Hin*dIII-digested λ DNA fragments.

Restriction endonuclease reactions

All reactions involving restriction endonucleases were performed under conditions recommended by the manufacturers. Samples of genomic DNA (10 µg) were restricted with 20 units of enzyme for at least 16 hours.

Polymerase chain reaction (PCR)

General PCR reactions were performed in a total volume of 25 µl, with the reaction mixture containing approximately 100 ng of template DNA; 0.4 mM each of dATP, dCTP, dGTP and dCTP; 1x PCR reaction buffer; 1.5 units *Taq* DNA Polymerase; 2 mM $MgCl_2$ and 100 ng of each primer. DNA amplification was performed in a programmable PTC-200 DNA Engine (MJ Research, Inc.). Typically, conditions involved an initial denaturation at 95°C for two minutes, followed by 25 cycles of denaturation at 95°C for one minute, annealing for one minute at 55°C and extension at 72°C for four minutes. If the PCR products were intended for use in cloning with pGEM-T Easy, a final cycle with an extension time of seven minutes was included.

Colony PCR

Colony PCR was performed by picking single bacterial colonies into a PCR reaction mix, with PCR conditions as described above except for an initial denaturation period of five minutes and a total of 30 cycles.

DNA precipitation using ethanol

DNA was precipitated by addition of 0.1 volumes of 3 M sodium-acetate, pH 5.2 and 2-2.5 volumes of cold redistilled ethanol. The DNA was left to precipitate at -20°C for 16 hours and was recovered by centrifugation at 12,000g for 25 minutes at 4°C. The pellet was washed with cold 70% (v/v) ethanol, dried under vacuum and resuspended in 1x TE.

DNA precipitation using isopropanol

In some cases, 0.1 volumes of 3 M sodium-acetate, pH 5.2, and 1 volume of isopropanol were used to precipitate DNA. The DNA was recovered by centrifugation at 12,000 g for 15 minutes and all traces of isopropanol were removed using a drawn-out Pasteur pipette. The resulting pellet was washed in 70% (v/v) ethanol, dried under vacuum and resuspended in 1x TE.

Gel electrophoresis

DNA was size fractionated in 0.7-2.0% agarose-TAE gels (Sambrook *et al.*, 1989). DNA fragments were stained with ethidium bromide and visualised with a Gel Doc 2000 Gel Documentation System (Bio-Rad).

RNA samples (10 µg each) were washed with 70% ethanol, dried in a vacuum centrifuge and dissolved in the following: 3.5 µl RNase-free water, 2 µl 10x MOPS, 3.5 µl formaldehyde, 10 µl deionised formamide and 1 µl of 10 mg/ml ethidium bromide. The samples were then denatured by incubation at 65°C for 10 minutes and mixed with 5 µl of 5x RNA tracking dye. RNA samples were electrophoresed in 1.2% (w/v) agarose gels containing 5% (v/v) formaldehyde and 1x MOPS buffer. The gel was run at 50 volts with 1x MOPS as the running buffer followed by visualisation with UV illumination.

General cloning and bacterial transformation

Plasmid vector DNA for use in ligation reactions was prepared by digestion with the appropriate restriction endonuclease(s), followed by either gel purification with the QIAquick Gel Extraction Kit or heat inactivation. If required, dephosphorylation with calf intestinal phosphatase (CIP) or shrimp alkaline phosphatase (SAP) was performed as per manufacture's instructions. When CIP treatment was used, a final gel purification step was performed to remove the CIP, while a heat inactivation step of 15 minutes at 65°C was performed after SAP treatment.

Insert DNA was prepared by digestion with the appropriate restriction endonuclease(s) or by PCR, followed by a gel purification step. Ligation reactions were performed overnight at 16°C in a total volume of between 10 and 20 µl and contained approximately 100 ng of vector DNA, 400 ng of insert DNA, 1x Ligation Buffer (Promega) and 3 Weiss units of T4 DNA Ligase. Transformations were performed as described (Inoue *et al.*, 1990) using 200 µl of competent cells and 50-100 ng plasmid DNA.

DNA sequencing

DNA was sequenced using the ABI Prism™ Dye Terminator Cycle Sequencing Ready Reaction Mix (Perkin-Elmer), except for the use of half of the usual amount of reaction mix. 400 – 800 ng of double-stranded DNA was used as template and approximately 100 ng or 18 pmol of primer was used. Reactions were performed using an MJ Research PTC-200 Peltier Thermal Cycler, with the following conditions: 26 cycles of 96°C for 30 seconds, 50°C for 15 seconds and 60°C for four minutes. Dye terminator gels were run and analysed by the Sequencing Centre at the IMVS (Adelaide, South Australia). Sequencing results were visualised with the Chromas v2.23 program (Technelysium).

Sequence analyses

Nucleotide and protein sequences were analysed using various programs in the Wisconsin GCG Sequence Analysis Software suite, Version eight and later (Accelrys, San Diego, CA; Genetics Computer Corporation, WI, 1984; (Devereux *et al.*, 1984). These programs were accessed online via the Australian National Genomic Information Service (ANGIS, http://www.angis.org.au). Database comparisons were performed using the BLAST algorithm (Altschul *et al.*, 1990, Altschul *et al.*, 1997) and sequence databases available online through the National Centre for Biotechnology Information http://www.ncbi.nlm.nih.gov.

Potential open reading frames and translational start and stop codons were identified within each cDNA using the FRAMES program (ANGIS, http://www.angis.org.au).

Bioinformatic promoter analysis

Promoters were analysed for the presence of *cis*-acting regulatory elements by comparison to the PLACE (plant *cis*-acting regulatory DNA elements) database (Higo *et al.*, 1999) available online from the National Institute of Agrobiological Sciences (http://www.dna.affrc.go.jp/PLACE).

Promoter sequences were analysed for novel *cis*-acting regulatory elements with the two dimensional "dot plot" visual display approach (Maizel and Lenk, 1981). Pairs of promoter sequences were first aligned using the COMPARE program (ANGIS), and output files from this program were used as input files for DOTPLOT program (ANGIS). A window size of seven and a stringency setting of seven were used for promoter analysis.

2.2.3: Nucleic acid transfer to membranes

Southern transfer

Following electrophoresis, DNA was transferred to a Hybond-N^+ nylon membrane with 0.4 M NaOH as the transfer buffer, as per manufacture's instruction. After 16 hours of DNA transfer, membranes were washed in 2x SSC.

Northern transfer

Following electrophoresis of RNA, gels were rinsed several times in DEPC H_2O, and then soaked in 0.05 M NaOH, 1x SSC for 20 minutes to partially hydrolyse the RNA. The formaldehyde was removed by subsequent soaking of the gel in 20x SSC for 45 minutes. The RNA was then transferred onto Hybond-N^+ following manufacturer's instructions and fixed onto the membrane by blotting onto Whatman 3MM paper soaked in 0.05 M NaOH for five minutes, followed by rinsing in 2x SSC and storage at 4^oC.

Plaque lifts

Plaque imprints were transferred onto Hybond N^+ following manufacturer's instructions. Prior to plaque lifting, plates were chilled at 4°C for approximately two hours to stabilise the top layer of agarose. The Hybond N^+ filters were then carefully placed on top of the agarose surface and left for one minute at room temperature. The filters were peeled off and the DNA within the λ bacteriophage particle imprints was denatured, neutralised and fixed onto the filters, according to Amersham instructions.

2.2.4: Nucleic acid detection with radioactive probes

Radioactive labelling of DNA probes

DNA probes were generated by PCR or restriction digest and purified from agarose gels using the QIAquick Gel Extraction Kit. 50-200 ng of probe was labelled by primer extension of random decamer oligonucleotides (Feinberg and Vogelstein, 1983) using the method of (Hodgson and Fisk, 1987) and [α-^{32}P] dATP. Unincorporated nucleotides were removed by spin column chromatography through Bio-Gel P60 resin using 100 µl G50-100 beads and 400 µl G100-200 beads layered into a 1.5 ml microcentrifuge tube punctured with a

fine needle at the conical point. Columns were packed and equilibrated with 100 µl stop buffer by centrifugation at 170 *g* for two minutes. Labelled DNA fragments were collected by centrifugation of the labelled samples, in stop buffer to a total volume of 100 µl, through the column at 170 *g* for two minutes.

Single-stranded cDNA probes labelling

To radioactively label single-stranded cDNA probes for differential screening, 2 to 4 µg of poly(A)$^+$ RNA was dissolved in 5 µl of DEPC-treated H_2O by incubation at 65^{o}C followed by incubation on ice for five minutes. After briefly microcentrifuging the sample, the following components were added in order: 1 µl DTT (10 mM), 20 units RNasin, 3.1 µl Tris-HCl pH 8.3 (400 mM), 1.2 µl KCl (625 mM), 0.5 µl $MgCl_2$ (500 mM), 2.4 µl oligo(dT) (2.4 µg/µl, Amersham), 1.3 µl of dATP, dTTP and dGTP (10 mM), 120 µCi [α-^{32}P]dCTP and 1 µl MLV reverse transcriptase (200 units/µl). The mix was incubated at 41^{o}C for 2 hours. Unincorporated nucleotides were removed with spin column chromatography as above.

Radioactive labelling of DNA ladder

1 kb Plus Ladder was end-labelled using 3 µg of *Eco*RI digested ladder, 30 µCi of [α-^{32}P] dATP and 30 units of Klenow enzyme as per manufacturer's instructions.

Prehybridisation conditions

All filters were prehybridised in a freshly made solution of hybridisation buffer at 65^{o}C for a minimum of two hours. Plaque lift filters were prehybridised for at least 16 hours. Nylon grids were used to separate filters when multiple filters were hybridised together. The volume of prehybridisation solution was approximately 0.2 ml/cm^2 of filter.

Probe denaturation and hybridisation

Radiolabelled probe was denatured by the addition of 34 µl 1 M NaOH and incubated at room temperature for 10 minutes. The solution was neutralised by the addition of 34 µl 1 M HCl and the denatured probe stored on ice. Filters were hybridised at a concentration of approximately 1x10^6 counts per minute per millilitre of hybridisation solution (to a maximum volume of 20 ml) for a minimum of 16 hours. Oligo(dA) was added to the cDNA library hybridisations at a concentration of 5 µg/ml of hybridisation solution. All hybridisations were performed at 65^{o}C overnight.

Colony hybridisation

Colony hybridisations were performed largely as described by Grunstein and Hogness, 1975. Following overnight colony regrowth at 37°C on supplemented L-Agar plates, the nylon filter was placed colony-side-up on a pad of three pieces of Whatman 3MM paper soaked in lysing solution for 20 minutes, with blotting onto paper towel every five minutes. After lysis, the filter was transferred to a fresh pad of 3MM paper soaked in denaturing solution plus 1% SDS for a total of 20 minutes, with blotting every seven minutes. After denaturation, the filter was transferred to a fresh pad of 3MM paper soaked in neutralising solution for a total of 20 minutes, with blotting every seven minutes. After neutralisation, the filter was partially air-dried and washed in 2x SSC, with the bacterial debris removed with absorbent cotton wool. The DNA was alkali-fixed onto the filters as per Amersham instructions.

Filter washes

Filters were washed sequentially in 2x SSC with 0.1% SDS, 1x SSC with 0.1% SDS and 0.5x SSC with 0.1% SDS at 65°C for 30 minutes each. If required, filters were subjected to a higher stringency wash in 0.1x SSC with 0.1% SDS at 65°C for varying lengths of time.

Autoradiography and filter stripping

Filters were covered in plastic wrap to prevent them from drying out and exposed to X-ray film at varying temperatures (-70^{o}C, -20^{o}C, 4^{o}C or room temperature) for an appropriate period of time, backed by an intensifying screen (DuPont Hi-Plus). Typically, Southerns were exposed for between several minutes and several hours at room temperature, genomic Southerns for at least seven days at -70^{o}C and Northerns for between 24 hours and several days, usually at -70^{o}C

Southern and library filters were stripped of probe by boiling for 10 minutes in 0.1x SSC, 1% SDS. Northern blots were stripped of probe by pouring boiling 0.1x SSC onto them and allowing the solution to cool to room temperature with mixing. All stripped filters were rinsed in 2x SSC and stored in plastic wrap at 4^{o}C.

2.2.4: <u>5 DPA boll wall library construction and screening</u>

Poly(A)$^{+}$ RNA preparation

Poly(A)$^{+}$ RNA from leaf and 5 DPA boll wall was affinity purified on oligo(dT)-cellulose columns (mRNA Purification Kit) according to the manufacturer's instructions.

cDNA synthesis

Size fractionated double-stranded boll wall cDNA was synthesised from poly(A)$^+$ boll wall RNA using an Invitrogen cDNA Synthesis Kit, according to manufacturer's instructions. cDNA synthesis was monitored by incorporation of 10 μCi of [α–^{32}P]dCTP during second strand synthesis, followed by electrophoresis of a sample on an alkaline agarose gel and subsequent autoradiography.

5 DPA boll wall cDNA library construction

Double-stranded 5 DPA boll wall cDNA was ligated into *Eco*RI-digested and dephosphorylated λZAPII vector in accordance with the protocol supplied with the cDNA Synthesis Kit. Phage DNA from the ligation reaction was packaged using Packagene *in vitro* Packaging Extracts and plated with *E. coli* XL1-Blue cells at a density of 1500 plaque forming units per 150 mm plate (total plated library size of ~12,000 pfu).

5 DPA boll wall cDNA library screening

Phage from the library plates were lifted in duplicate onto Hybond-N$^+$ membranes to create two duplicate library filter sets. The filters were differentially screened with ^{32}P-labelled single-stranded cDNA probes from leaf or boll wall tissues. 30 plaques that more strongly hybridised the boll wall probe than the leaf probe on the primary screen were eluted by plugging the plaque into 0.5 ml SM buffer with 20 μl chloroform with overnight 4^{o}C incubation. Phage were replated at low density on 85 mm plates for secondary screening. Following secondary screening, recombinant pBluescript SK(-) were rescued from λZAPII using the ExAssist/SOLR System, following manufacturer's instructions.

Identification of redundant clones

To identify clones that contained redundant cDNAs, all 30 phagemids were digested with *Eco*RI to drop out the insert, separated by gel electrophoresis and transferred onto Hybond-N$^+$. Each of the 30 cDNA inserts was radioactively labelled and sequentially used to hybridise the cDNA inserts on the membrane. Redundant clones were identified and the largest member of each group further analysed.

Reverse Northern analysis

As above, all 30 pBluescript phagemids were digested with *Eco*RI, separated by gel electrophoresis and transferred to Hybond-N$^+$. The membranes were hybridised with labelled poly(A)$^+$ RNA from leaf and boll wall tissues as the probe to determine relative boll wall and leaf expression for each clone.

cDNA abundance

Approximate cDNA abundance was determined by hybridisation of the boll wall cDNA library filters with labelled cDNAs. The number of cDNAs within the library that hybridise with each probe was then used as an approximate estimate of promoter strength.

2.2.5: Promoter identification and isolation

Genomic Southerns

Genomic Southerns were performed as per Sambrook *et al*. (1989), with 20 units of various restriction endonucleases used to digest 10 µg aliquots of genomic DNA for at least 16 hours.

Genome Walker® PCR

Identification of promoter sequences was achieved using the Universal Genome Walker® Kit following the manufacturer's instructions. Firstly, cotton genomic DNA was digested with 10 different restriction enzymes that produce blunt ended DNA fragments, five of which were included in the kit (*Dra*I, *Eco*RV, *Pvu*II, *Sca*I and *Stu*I) and five of which (*Sma*I, *Ssp*I, *Hae*III, *Alu*I and *Rsa*I) were added for this study. Genome Walker® adapters were ligated onto the ends of the genomic fragments, in essence generating ten genomic "libraries". Following this, PCR reactions using gene-specific primers and primers specific to the adapters were undertaken. PCRs were performed using a polymerase mix of *Taq*/*Pfu* at a molar ratio of 16:1 in the supplied *Pfu* buffer.

Promoter isolation and cloning

Typically, the promoters used in this study were isolated from genomic DNA by PCR using primers containing appropriate restriction endonuclease recognition sites to allow cloning into the reporter vector pJK.KiwiGUS.ocs.Kmf(-). The promoters were first cloned into either pGEM-T Easy or pMDC162GUS and then transferred into pJK.KiwiGUS.ocs.Kmf(-). The *35S* positive control promoter was isolated from pMDC43GUS (Curtis and Grossniklaus, 2003) using the same technique.

2.2.6: Transient transformation

The particle injection gun

A custom-made particle injection gun (Vain *et al.*, 1993) was used for delivery of DNA-coated gold particles into plant tissues. The chamber and output areas of the particle injection gun were sterilised before use with 70% ethanol and the entire apparatus was treated

with UV light for approximately five minutes. A helium output pressure of 90 or 120 psi, an output release time of 0.12 milliseconds and a vacuum pressure of –90 kPa were the standard conditions used for biolistics. The distance at which the target tissue is located relative to the point of particle injection can also be adjusted, such that the tissue is situated at 25 mm intervals between 60 mm and 285 mm from the point of injection.

Preparation of gold particles

Gold particles (1.0 µm) were treated using a modified method from Franks *et al*. (1998). Approximately 8 mg of gold particles were thoroughly resuspended in 100 µl of 100% ethanol by vigorous vortexing for two minutes. After vortexing, the gold suspension was spun down for 10 seconds in a mini-centrifuge and the ethanol was removed. The gold particles were then resuspended in 100 µl of sterile dH_2O by pipetting up and down to avoid clumping of projectiles. After resuspension, the gold particles were spun down for 10 seconds in a mini-centrifuge. This process was repeated several times (up to ten times) until the gold projectiles were clearly dispersed on the side of the tube after the 10 second spin. Following the washes, the gold particles were resuspended in 100 µl of dH_2O.

Preparation of Swinney filters

Before use, each Swinney filter was thoroughly scrubbed and washed in 1 M HCl, followed by several washes in dH_2O, then autoclaved. Immediately before use, each filter was sterilised in 100% ethanol for approximately 15 minutes. Filters were air-dried and treated with UV light for five minutes prior to desiccation.

Precipitation of DNA onto gold particles

Using a method based on Franks *et al*. (1998), 8 mg of washed gold particles were thoroughly resuspended in 100 µl of dH_2O. 10 µl of plasmid (1 µg/µl) was slowly added to the resuspended gold particles, followed by the addition of 40 µl of 1 M spermidine while vortexing at medium speed. 100 µl of 2.5 M $CaCl_2$ was then added to the suspension while vortexing. To facilitate precipitation of the plasmid onto the gold particles, the mixture was incubated on ice for 10 minutes. A 10 second centrifugation step was then used to pellet the plasmid-coated gold particles followed by resuspension in 110 µl of 100% ethanol. 12 µl of resuspended DNA-coated gold particles were added to each of eight sterile Swinney filters.

Target tissue preparation

Plant tissues used for particle bombardment were removed from the plant and placed on ice. Individual tissue sections were isolated under sterile conditions using a scalpel blade

and placed onto ½ MS plates containing chloramphenicol (50 μg/ml). Young leaves were simply removed from the plant while bract, calyx and petal tissues were isolated as described above. Fibres were isolated by removing the bracts and calyx tissues from a boll, followed by cutting open the boll wall along the carpel wall lines to expose the locules. Locules, containing fibres and ovules, were isolated and stored on ice until required. Boll wall tissues were isolated by cutting thin (<1 mm) sections along the side of the boll as shown in figure 2.1 and were stored on ½ MS plates until transformation.

Particle bombardment

Plant tissues for transformation were placed on sterilised cardboard and covered with a mesh grid, which was fixed to the cardboard by drawing pins. The cardboard was then placed into the bombardment chamber at the appropriate level for the tissue to be bombarded. 0 to 30 DPA boll wall, 1 DPA bract, young leaf and 3 to 4 DPA fibre tissues were bombarded at a distance of 110 mm from the particle injection point. Petals were bombarded at a distance of 135 mm and 13 DPA calyx tissue at a distance of 185 mm. A helium output pressure of 90 psi was used for all tissues except boll wall, for which 120 psi was required in order to obtain sufficient transformation efficiencies. After bombardment, tissues were placed onto ½ MS agar plates containing 50 μg/ml chloramphenicol and placed under fluorescent light at 25°C overnight, except for fibre tissue which was shielded from the light.

Histochemical staining of bombarded tissues

Subsequent to the overnight incubation, bombarded tissues were assayed for GUS activity using a method adapted from Franks *et al*. (1998). Tissues were placed into appropriately sized tubes and prefixing solution was added to submerge the sample. The solution was vacuum infiltrated for 10 minutes. The prefixing solution was removed and the tissues were washed three times in 0.1 M $NaPO_4$ buffer. After washing, assay buffer was added to tissues and the samples were incubated overnight at 37°C. Following overnight incubation, chlorophyll was removed from the tissues by incubation in 100% ethanol for five minutes at 65°C. Tissue samples were then stored in 70% ethanol. Histochemically stained tissues were viewed and photographed with SMZ800 Nikon light microscopes and photographed with a Coolpix digital camera (Nikon, Japan). Photoshop 6.0 was used for image preparation.

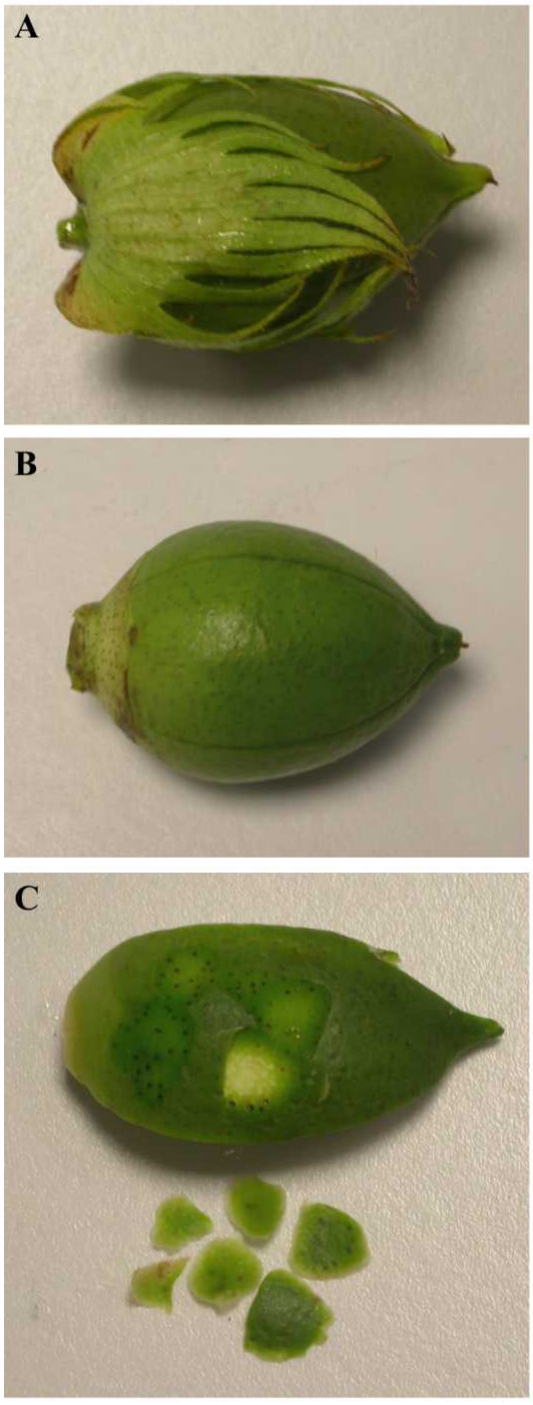


Figure 2.1: Preparation of boll wall tissue for transformation.
Bolls were isolated from the plant (A), bracts were then removed (B), followed by removal of calyx tissue and slicing of thin sections for transient transformation (C).

Chapter 3: Isolation and characterisation of cotton boll wall-specific mRNAs

3.1: Introduction

To identify cotton promoters capable of driving boll wall-specific expression, mRNA transcripts that are more abundant in boll wall than in leaf cells were identified by differential screening. The promoter regions corresponding to these transcripts were then isolated. The method of differential screening allows the abundance of individual mRNAs within a cDNA library to be compared between two tissues or timepoints. This type of screen selects for mRNA transcripts of greater than 0.05% abundance in the target tissue (boll wall) and less than 0.01% abundance in the second tissue (Sambrook *et al.*, 1989), which is ideal for this study as it selects for highly abundant transcripts and hence strong promoters. 5 DPA boll wall was selected because this timepoint occurs during the peak period of boll wall expansion between 1 DPA and 15 DPA (Wilkins and Jernstedt, 1999) and it is a time when the boll is most vulnerable to biotic and abiotic stress. The tissue selected for comparison with boll wall was leaf because this tissue represents the main biomass of the plant and is the tissue from which it is most desirable to eliminate transgene expression. Young expanding leaves were used because they are more transcriptionally active than older, mature leaves.

3.2: Isolation of putative boll wall-specific cDNA clones

Cotton tissues are not amenable to standard plant RNA extraction methods as they contain inhibitory molecules such as phenolics, terpenes, polysaccharides and secondary metabolites (Katterman and Shattuck, 1983, Galau *et al.*, 1988, Baker *et al.*, 1990, Schneiderbauer *et al.*, 1991, John, 1992). Therefore, the labour-intensive method of Wan and Wilkins (1994) was used for RNA extraction. This method utilises substances such as PVP, spermidine and proteinase K, as well as specific buffers and long incubation periods to overcome the presence of inhibitory molecules.

RNA was prepared from boll wall and leaf tissues followed by poly$(A)^+$ RNA isolation. cDNA was then generated from the boll wall poly$(A)^+$ RNA, with synthesis monitored by incorporation of [$\alpha-^{32}$P] dCTP during second strand synthesis. The technique utilised to prepare the cDNA involved a size fractionation step, to select longer cDNAs that are more likely to be complete. The cDNA was viewed by electrophoresis on an alkaline agarose gel and subsequent autoradiography (figure 3.1). The autoradiograph showed a prominent band at about 2,300 bp, overlaying a smear of cDNA products ranging from 1,000 bp to 4,500 bp.

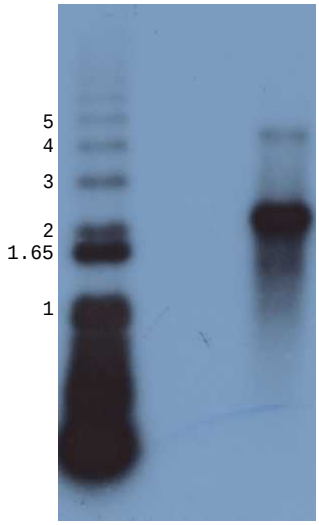


Figure 3.1: Autoradiograph of cDNA products from second strand synthesis.
cDNA was prepared for library construction from 5 DPA boll wall poly(A)$^+$ RNA. A portion of the cDNA was labeled with [$\alpha-^{32}$P] dCTP and separated on an alkaline agarose gel, followed by autoradiography. The majority of the cDNA is of approximately 2.3 kb in size with products visible between 1 kb and 4.5 kb. Marker lane sizes are indicated in kb.

The size-fractionated cDNA was cloned into λZAPII generating a library of 12,000 plaque-forming units or clones. By using conventional +/- screening (Sargent, 1987) with boll wall and leaf poly(A)$^+$ RNA probes, 30 plaques with the strongest boll wall hybridisation and the weakest leaf hybridisation were selected for further analysis (figure 3.2). The hybridisation patterns for each of the 30 clones were reassessed after clone isolation by secondary screening. All clones showed preferential hybridisation to the boll wall probe (figure 3.3).

To determine cDNA sizes, restriction digestion and gel electrophoresis was used to separate the cDNA inserts from their plasmids. Insert sizes range from 225 bp to 4,300 bp with an average of 1,517 bp (figure 3.4). The average cDNA size before the library was cloned (2,300 bp) and the average size of inserts obtained from the differential screen (1,517 bp) indicates that the library reflected relatively long transcripts. Reverse Northern analysis was used to eliminate the clones that contained cDNAs that hybridised weakly with labelled poly(A)$^+$ boll wall RNA or relatively strongly with labelled poly(A)$^+$ leaf RNA (figure 3.5). To identify duplicate or redundant cDNA clones, cross hybridisation experiments were performed (figure 3.6). Following the elimination of clones that were redundant or corresponded to transcripts transcribed in undesirable spatial patterns, eight distinct clone types remained. The largest clone within each class of cDNA was sequenced and compared to sequence databases to establish putative identity. The clone groups are named after these representative clones (figure 3.6) and for the duration of this thesis, the cDNAs and promoters will be referred to by these names. Four of the eight clone types are unique within the 30 clones selected from the library and show sequence similarity to δ-tonoplast intrinsic proteins (δ-TIP), actin, dihydroflavonol 4-reductase (DFR) and anthocyanidin reductase (ANR). Four clones were identified multiple times within the library; six clones corresponding to a proline rich protein (PRP), five clones corresponding to chalcone synthase (CHS), four clones corresponding to sucrose synthase (SuSy) and three clones corresponding to granule-bound starch synthase (GBSS).

3.3: Temporal and spatial expression patterns of putative boll wall-specific mRNA transcripts

To investigate the transcription patterns of the putative boll wall specific mRNAs in tissues other than the boll wall and leaf, Northern analysis of several tissues was performed. From the Northern analysis it was evident that the δ-TIP and actin mRNA were present at significant levels in tissues other than the boll wall. δ-TIP mRNA was detected at high levels in stem and flower tissues and at low levels in leaf and root (data not shown). Actin mRNA

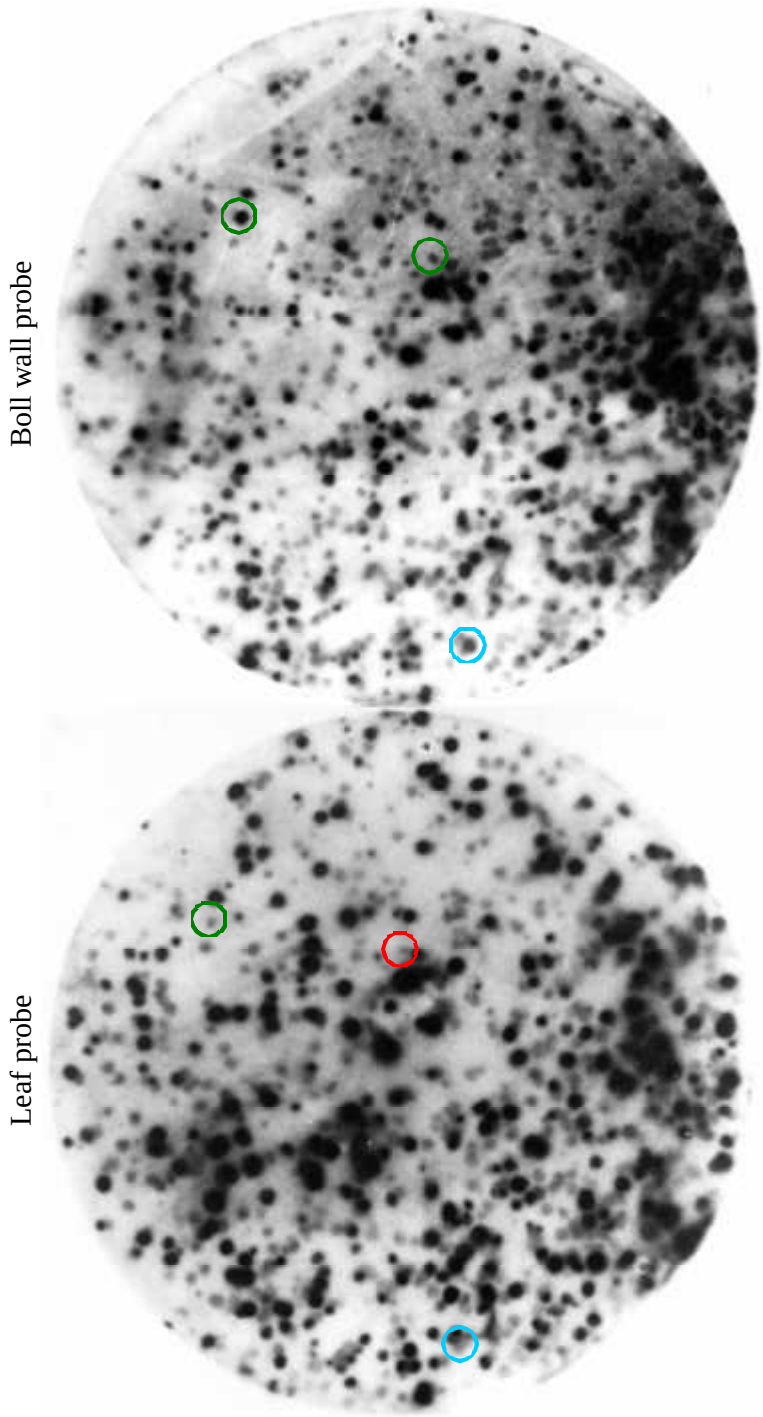


Figure 3.2: Primary differential screen.
Hybridisation to one of the plaque lifts of the library is shown as an example of the differential screen. The top autoradiograph shows hybridisation to the boll wall cDNA probe and the bottom autoradiograph shows hybridisation to the leaf cDNA probe. Several differentially hybridising clones are indicated with circles and correspond to anthocyanidin reductase (green), a proline rich protein (blue) and chalcone synthase (red).

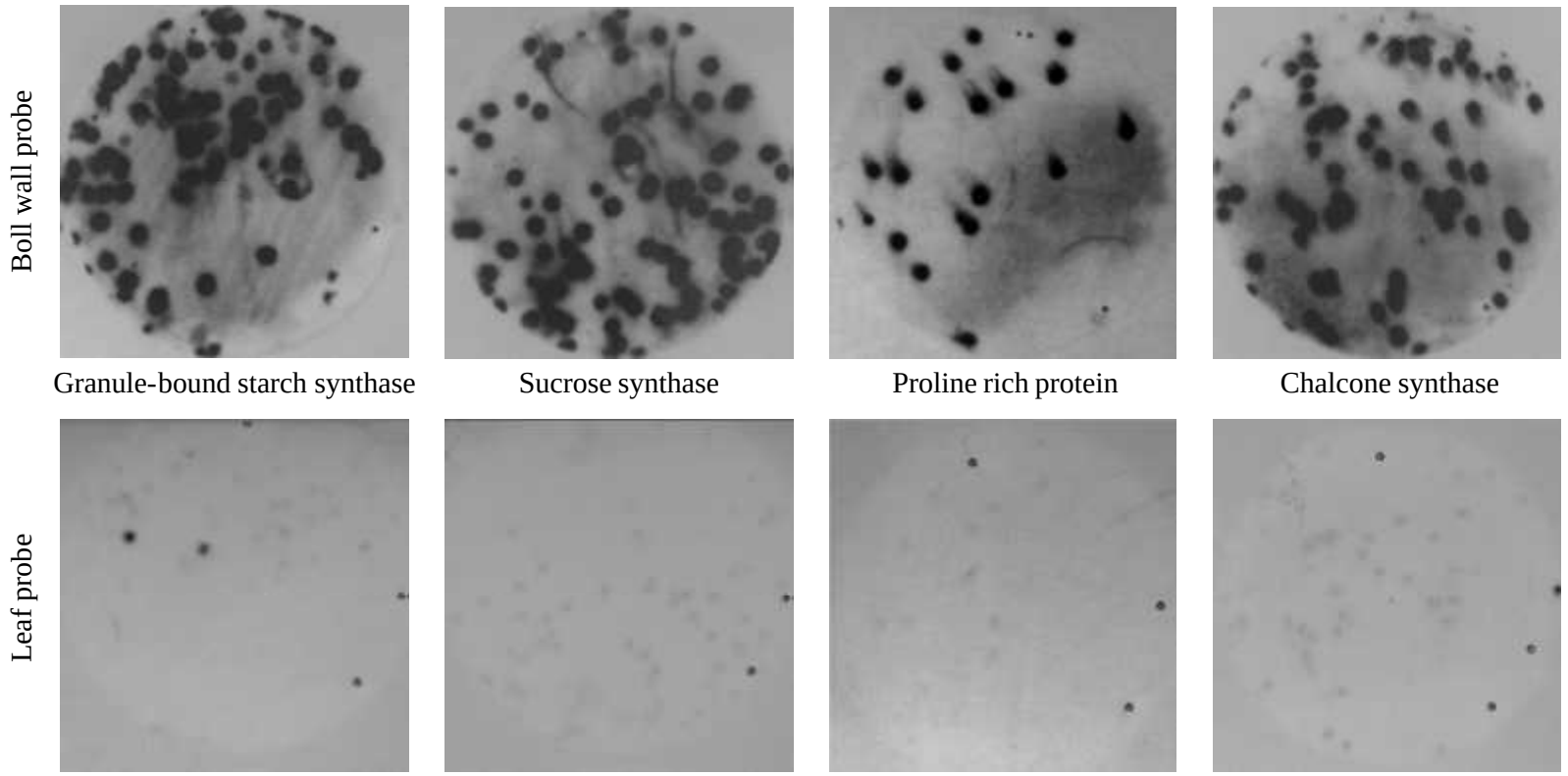
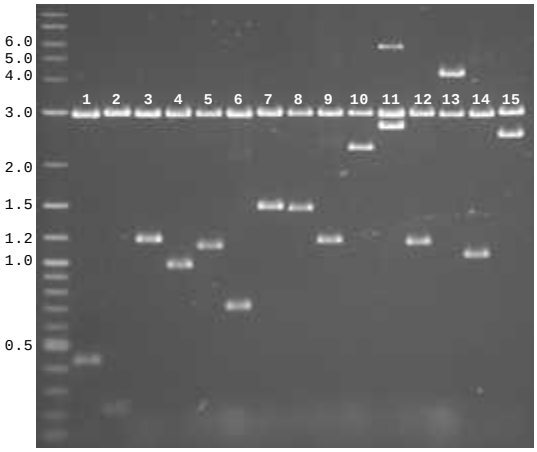
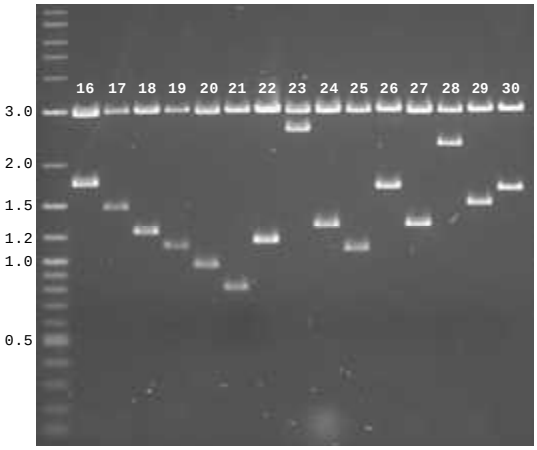


Figure 3.3: Secondary screening of putative boll wall-specific clones.
Autoradiographs of the granule-bound starch synthase, sucrose synthase, proline rich protein and chalcone synthase plaque lifts are shown after hybridisation with labelled 5 DPA boll wall cDNA and leaf cDNA, as indicated.





Clone	Identity	Size (bp)
1	Proline rich protein	450
2	Starch synthase	225
3	Sucrose synthase	1200
4	N/A	975
5	Proline rich protein	1100
6	Chalcone synthase	700
7	Chalcone synthase	1475
8	Chalcone synthase	1450
9	Proline rich protein	1125
10	Starch synthase	2300
11	Sucrose synthase	2900
12	Proline rich protein	1125
13	N/A	4300
14	Tonoplast intrinsic protein	1025
15	N/A	2600
16	N/A	1800
17	Chalcone synthase	1500
18	Sucrose synthase	1275
19	Proline rich protein	1150
20	N/A	1000
21	N/A	825
22	Proline rich protein	1200
23	Sucrose synthase	2800
24	Anthocyanidin reductase	1300
25	N/A	1175
26	Actin	1700
27	Dihydroflavonol reductase	1300
28	Starch synthase	2300
29	Chalcone synthase	1525
30	N/A	1700

Figure 3.4: cDNA sizes.
Restrictions digests with either *Eco*RI or *Not*I were performed on the 30 clones (numbered 1 to 30) to excise the cDNA inserts from the phagemid vectors. The products were then separated on agarose gels as shown. The sizes of the cDNA inserts were determined by comparison to the marker lanes and are shown in the table. The identity of each of the clones, as assigned by comparison to sequence databases, is also shown in the table. Marker lane sizes are indicated in kb.

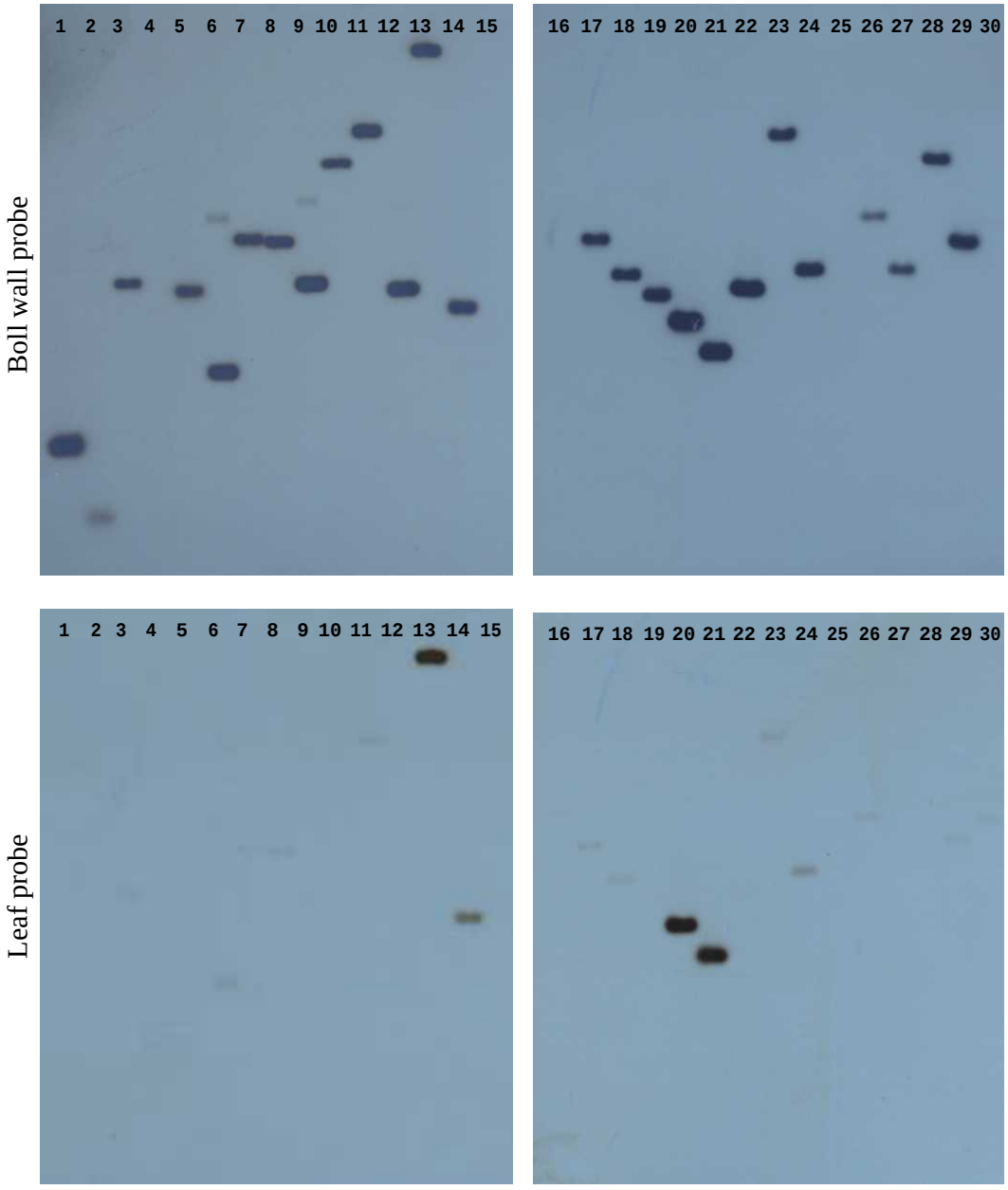


Figure 3.5: Reverse Northern analysis of putative boll wall specific clones.
Digestion and gel electrophoresis of the 30 clones was performed followed by transfer to membranes. The membranes were hybridised with total RNA from 5 DPA boll wall (top) and leaf (bottom) tissues. The transcripts corresponding to several clones (#4, #15, #16, #25 and #30) were rare in the boll wall, while the transcripts corresponding to several other clones (#13, #14, #20 and #21) were relatively abundant in the leaf tissue.

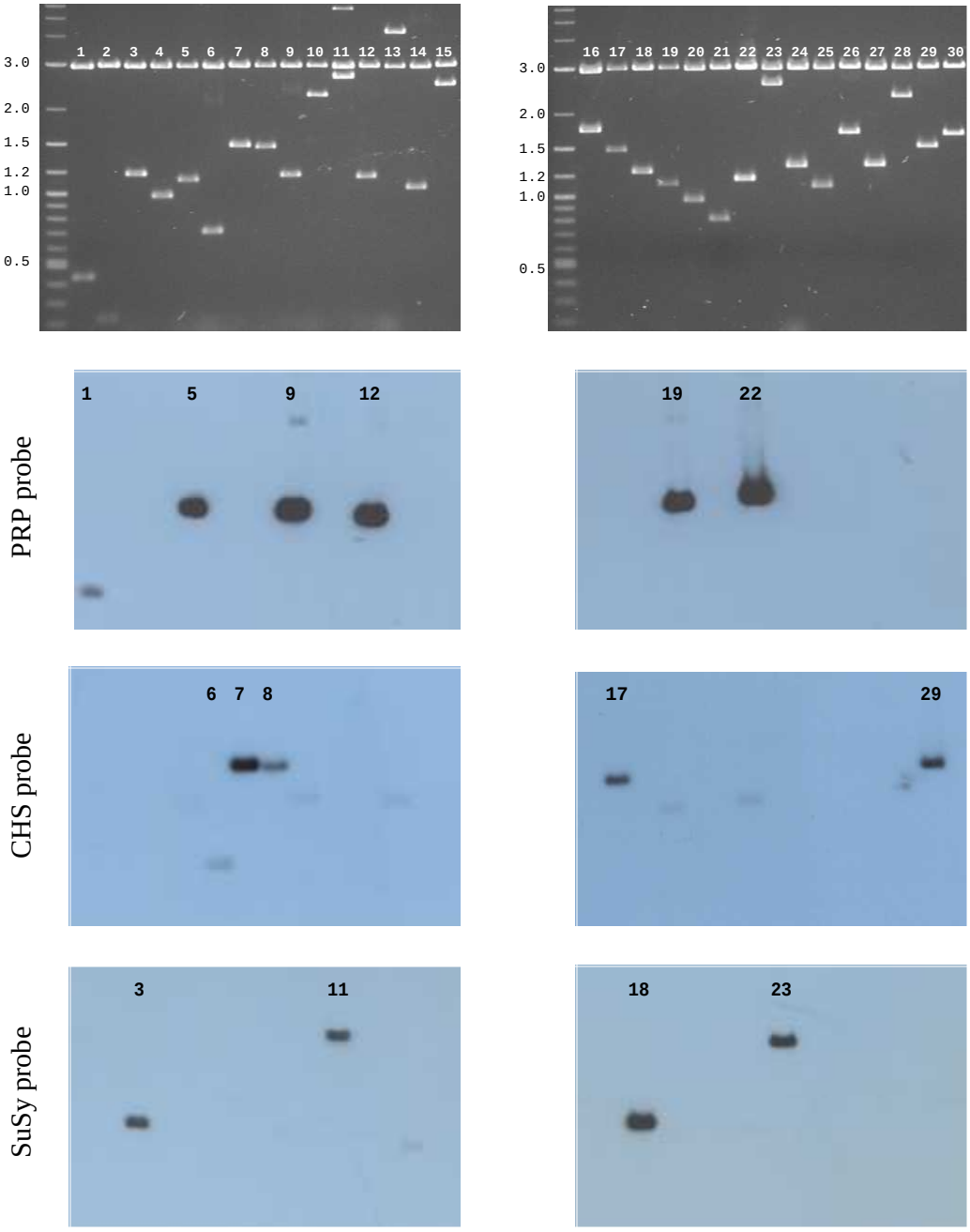


Figure 3.6: Detection of redundant clones.
Clones were digested and transferred to membranes as previously discussed. The membranes were hybridised with individual labelled cDNAs to identify clones that were isolated multiple times in the original screen. Shown above are examples of the autoradiographs of the filters hybridised with the PRP, CHS and SuSy cDNAs as indicated. Marker lane sizes are shown in kb.

was moderately abundant in most tissues tested (data not shown). While the abundance of these two transcripts is relatively low in leaf tissue, consistent with their detection in the screen, the promoters are unsuitable for this study because they drive significant expression in non-target tissues.

To further investigate the temporal and spatial transcription patterns of the mRNA transcripts corresponding to the six remaining cDNAs, a more comprehensive Northern analysis was performed (figure 3.7). The levels of the six mRNAs were assessed in a variety of tissue types and during development of the boll wall. In general, all six mRNAs are present at relatively high levels during the early stages of boll wall development, with decreasing abundance as the boll matures. The transcripts are generally most abundant during the period of maximal boll enlargement between 1 and 15 DPA (Wilkins and Jernstedt, 1999), with diminishing abundance thereafter. As expected, the transcripts were highly abundant in 5 DPA boll wall, which is the tissue of origin of the library. Also as expected, mRNA levels in the leaf are either low or non-existent for these transcripts. For several of the clones, some minor levels of transcript can be seen in tissues other than the boll wall, such as fibre and petal.

The PRP mRNA was detected in boll wall tissues from 0 to 25 DPA (figure 3.7). While a low level of transcript was detected at 0 DPA, higher levels were present at 5 DPA, with declining abundance from 10 to 25 DPA. No mRNA is evident in other tissues, suggesting that the expression of this gene may be boll wall-specific. The CHS transcript is detectable in the boll wall between 0 and 25 DPA, with high levels between 0 and 15 DPA. Low abundance is seen in the fibre and root tissues and a high level is observed in the petal tissue. The most persistent boll wall expression is seen for GBSS, with mRNA still detectable at 40 DPA. High levels of GBSS mRNA are seen in the boll wall from 0 to 25 DPA, declining to become undetectable by 50 DPA. Minor levels are seen in several tissues including leaf, fibre, bract and calyx. The most widely expressed mRNA is that encoding SuSy, with expression detected in most of the tissues analysed. The pattern of boll wall expression for this transcript is similar to that of GBSS, with high levels between 0 and 10 DPA declining over the next 15 days. The mRNAs encoding DFR and ANR exhibit similar patterns to each other with low levels of DFR mRNA detected in the boll wall between 0 and 10 DPA and low levels of ANR mRNA detected in the boll wall between 0 and 15 DPA. Low levels of both transcripts were also present in fibres.

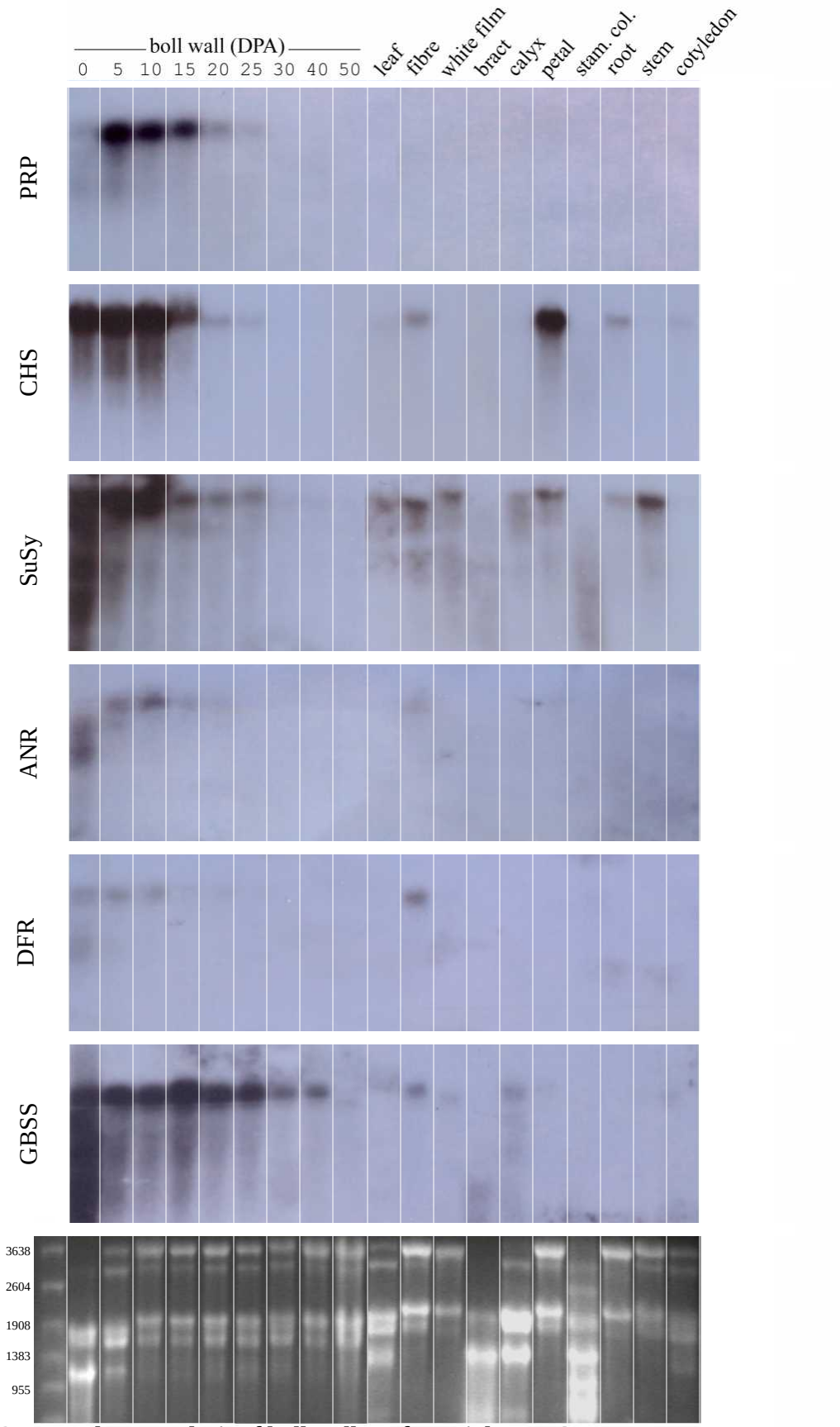


Figure 3.7: Northern analysis of boll wall-preferential mRNAs.
The abundance of the six mRNA transcripts in various tissues was assessed by Northern analysis. 10µg of each RNA was electrophoresed, transferred to a membrane and hybridised with each of the cDNAs. The bottom panel shows a representative ethidium bromide stained RNA gel used to monitor RNA loading. Marker lane sizes are indicated in nucleotides.

The cDNAs that were identified multiple times in the library screen (those corresponding to PRP, GBSS, CHS, and SuSy) appear to be the most highly abundant in the boll wall at 5 DPA, and all have detectable levels of mRNA transcript persisting later in boll wall development. The two clones that were identified only once within the clone population were of low abundance at 5 DPA in boll wall tissue, with their mRNAs detectable in Northern blots until only 10 or 15 DPA.

As can be seen in figure 3.7, ethidium bromide-stained RNA was used to monitor RNA loading. The RNA that was prepared from 0 DPA boll wall and bract tissues has a lower abundance of the 28S ribosomal RNA band, at approximately 3,600 nucleotides, than the other RNA samples suggesting degradation of the RNA has occurred in these samples. Repeated and more thorough preparation of RNA from these tissues was performed without any improvement in RNA quality. The methods used for RNA extraction from cotton are poorly developed and preparation of RNA is usually limited to tissues such as leaf and fibre, with no published reports of RNA extraction from boll wall or bract tissues. Presumably, the inhibitory compounds in cotton tissues described previously, which are known to interfere with RNA extraction, are the cause of the problem. However, faint 28S ribosomal RNA bands can still be seen, and upon hybridisation with a probe that is complementary to mRNA transcripts in the 0 DPA boll wall sample, hybridisation to a band of similar size as that seen in the other RNA samples is apparent with little evidence of degradation (figure 3.7). This observation suggests that long transcripts are present in these samples in sufficient quantities to allow reliable estimation of mRNA abundance.

3.4: Sequence analysis of boll wall-specific cDNA

Open reading frames for each of the cDNAs were identified using the FRAMES program (ANGIS, http://www.angis.org.au) to identify putative start and stop codons within the cDNAs and hence, potential open reading frames (appendices 1A-H). In each cDNA, one frame exhibited a single, major open reading frame. In addition, the reading frame of the cDNA sequences was compared to the reading frame of similar, previously characterised sequences within sequence databases.

3.4.1: δ-tonoplast intrinsic protein

One of the unique clones showed 99.2% nucleotide similarity, including 100% identity within the coding region, to a previously-identified gene encoding cotton δ-TIP (Ferguson *et al.*, 1997), as well as approximately 84% amino acid identity to δ-TIPs from several plant species including *Nicotiana glauca* (Smart *et al.*, 2001) and *Arabidopsis*

(GenBank accession number NP_188245). The δ-TIP identified here is 987 bp long and contains 42 bp and 198 bp respectively of 5' and 3' putative untranslated regions (UTRs) (figure 3.8). A potential polyadenylation signal (AATAAA) is located within the 3' UTR, with similarity to the consensus polyadenylation sequence of AAATAA (Joshi, 1987a). The 747 bp open reading frame conceptually encodes a protein of 248 residues with a predicted molecular mass of 25 kDa and an isoelectric point of 5.5. The putative amino acid sequence includes conserved regions commonly found in major intrinsic proteins and a conserved cysteine residue that is reported to bind mercury in *Arabidopsis* δ-TIPs (Ludevid *et al.*, 1992, Ferguson *et al.*, 1997). Major intrinsic proteins are a large family of channel forming proteins that are found in all organisms (Pao *et al.*, 1991) that act as membrane channels for water (John, 1992). Within the plant kingdom, these channel-forming proteins are known as TIPs (Johnson *et al.*, 1990). TIP isoforms are often expressed during specific stages of plant development as well as in specific cell types (Ferguson *et al.*, 1997), with γ- and δ-TIPs found in elongating cells (Maeshima, 1990, Ludevid *et al.*, 1992, Maeshima, 1992, Maeshima *et al.*, 1994, Daniels *et al.*, 1996). TIPs have been implicated in cell elongation, including fibre expansion, and in stomatal opening and closure via water flux across guard cell membranes (Ferguson *et al.*, 1997, Smart *et al.*, 1998). While the level of the identified δ-TIP mRNA is higher in boll wall tissue than in leaf tissue, lower levels were seen in other tissue types including leaf, stem, root and flower, indicating a global function and limiting the usefulness of the corresponding promoter to this study.

3.4.2: Actin

Another clone that was unique amongst the clone set showed homology with previously identified cDNAs that encode actin. The clone identified here is 1,654 nucleotides long and conceptually encodes a protein of 377 amino acids, with a mass of 42 kDa and an isoelectric point of 5.3. The cDNA contains a 126 bp 5' UTR, a 394 bp 3' UTR and several potential polyadenylation signals (figure 3.9). The sequence identified here is similar to several cDNAs identified by Li *et al*. (2005) showing at least 99% nucleotide identity to two cDNAs, ACT3 and ACT13, and 100% amino acid identity with ACT2. Most of the 15 clones identified by Li *et al*. had relatively low abundance in leaf tissue compared to other tissues tested, as demonstrated by real time PCR, which accounts for the identification of actin cDNAs in the screen described here.

In plants, the actin cytoskeleton plays an important role in cell morphogenesis (Mathur *et al.*, 1999, Kost and Chua, 2002) and is essential for cell elongation and tip growth, possibly due to its role in the transportation of organelles and vesicles carrying membranes and cell

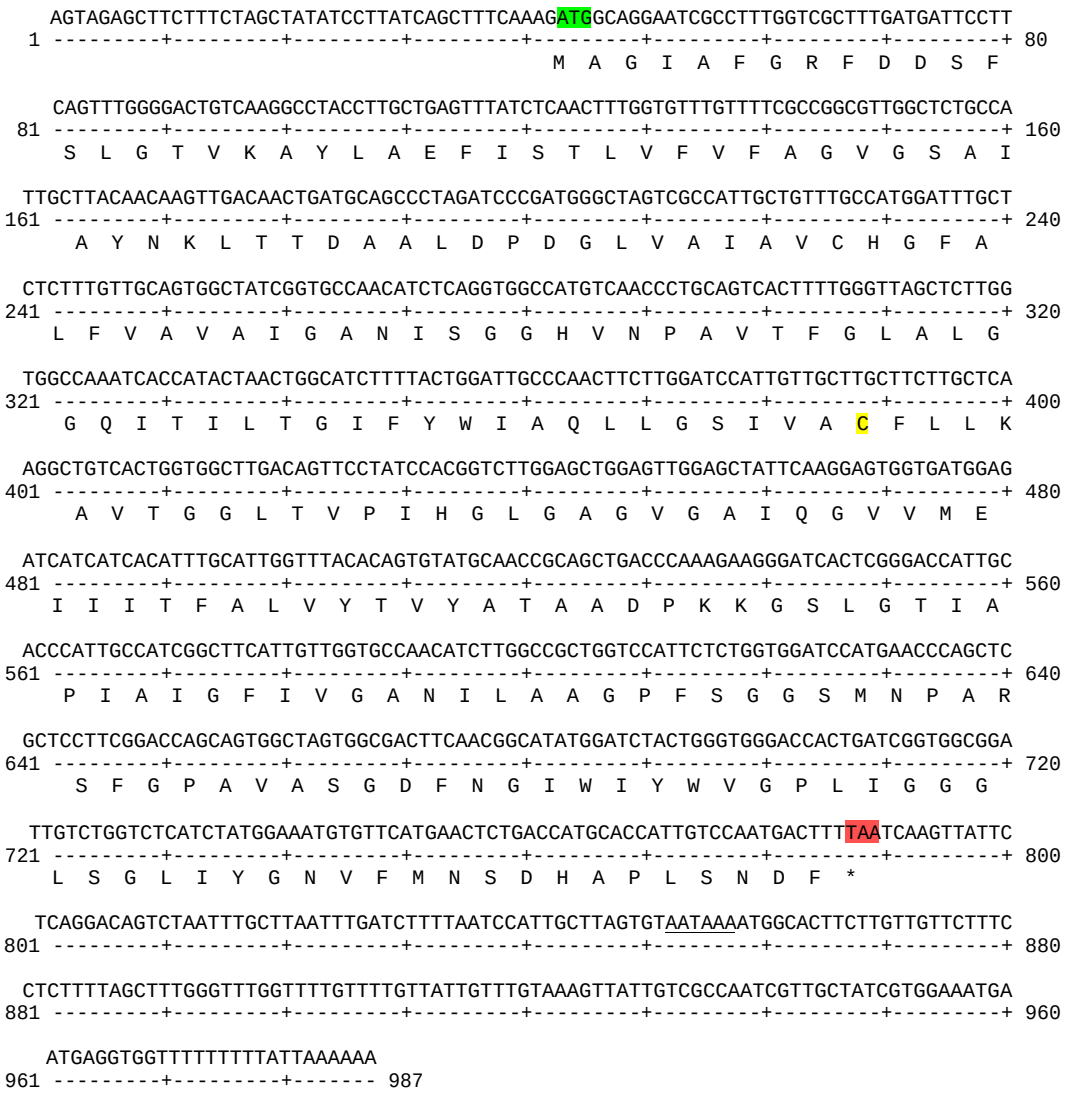


Figure 3.8: Nucleotide and conceptual amino acid sequences of δ-tonoplast intrinsic protein: Nucleotide and conceptual amino acid sequences of the isolated δ-TIP cDNA, showing the putative translation start site (green shading) and the putative translation termination codon (red shading). The cysteine residue that is important for mercury binding in *Arabidopsis* δ-TIPs (Ludevid *et al.*, 1992, Ferguson *et al.*, 1997) is shaded yellow. A potential polyadenylation signal site is underlined in black.

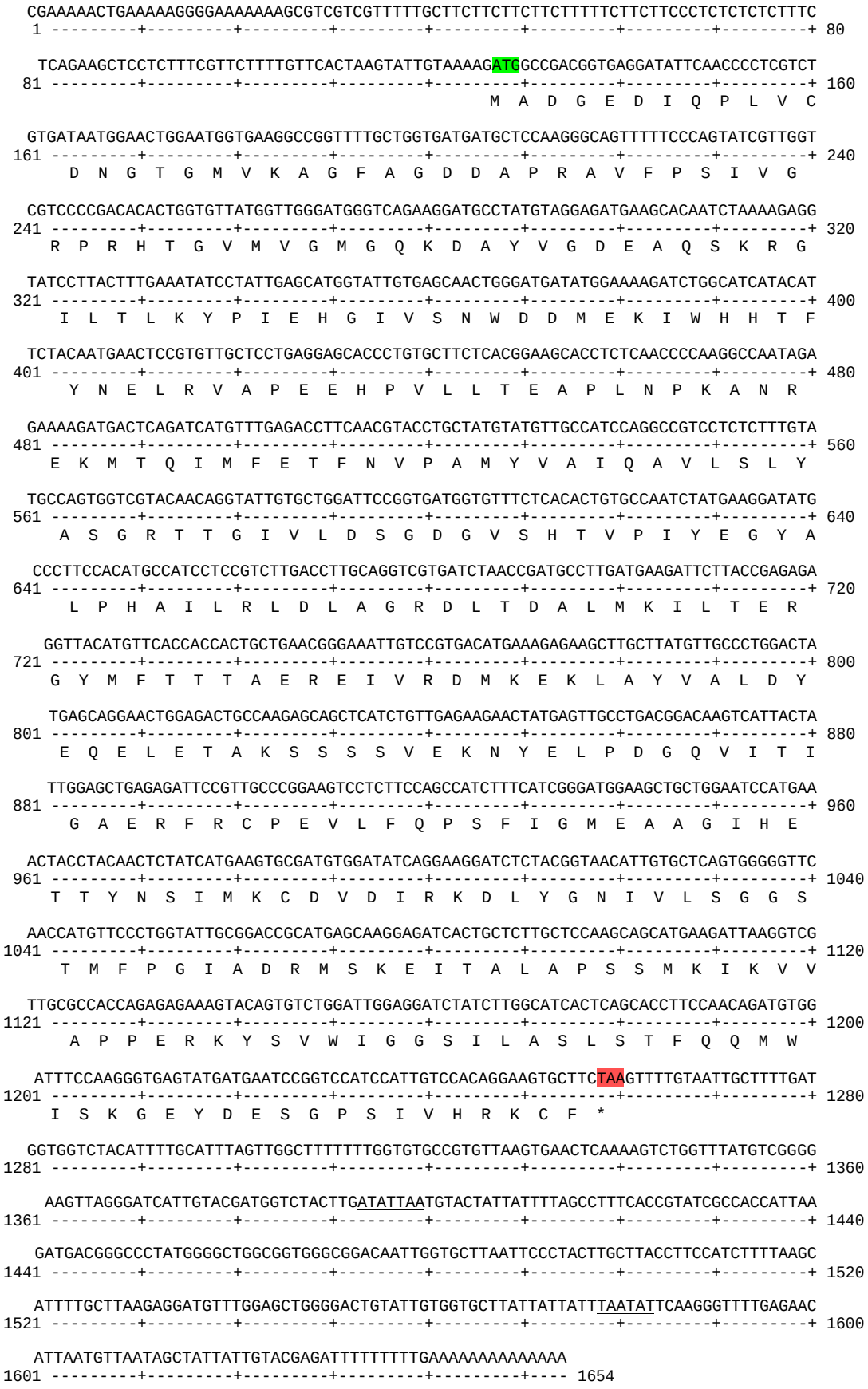


Figure 3.9: Nucleotide and conceptual amino acid sequences of actin.
Nucleotide and conceptual amino acid sequences of the isolated actin cDNA, showing the putative translation start site (green shading) and the putative translation termination codon (red shading). Two potential polyadenylation signal sites are underlined in black.

wall components to the site of cell growth (Li *et al.*, 2005). Therefore expression of actin in the boll wall would not be unexpected, as the boll enlarges rapidly in response to the expanding fibres inside.

3.4.3: Proline rich protein

Six of the clones identified were of the PRP class. The longest of the clones is 1,105 bp long, containing an 873 bp open reading frame with a 17 bp 5' UTR, a 215 bp 3' UTR and a potential polyadenylation signal (AATAAA) 22 bp before the poly(A)$^+$ tail (figure 3.10). The cDNA conceptually encodes a protein that contains an N-terminal signal sequence that is predicted to be cleaved between residues 29 and 30, most likely when the protein is exported out of the cell. After cleavage, the conceptual protein is 261 amino acids in length with a mass of 28 kDa, an isoelectric point of 10.7 and is composed of 27% proline residues. There are also ten semi-conserved repeats within the N-terminal region of the type (P/K/Q)-P-P-(V/T)-(H/Q/K)-P-P-(P/T/S) and a histidine rich region between residues 23 and 43 where 12 of 20 residues are histidine, including a six-histidine tract. The conceptual protein contains one putative N-glycosylation site (Asn-X-Ser/Thr) and an N-terminal PAC domain.

The conceptual protein shows moderate amino acid homology with several predicted plant PRPs: 59% homology with PvPRP1 from bean (Sheng *et al.*, 1991), 58% homology with DcAGP1 from carrot (Baldwin *et al.*, 2001), 48% homology with CaPRP1 from capsicum (Mang *et al.*, 2004) and 46% homology with NaPRP4 from Tobacco (Chen *et al.*, 1993). *PvPRP1* encodes a proline-rich cell wall protein that is down regulated in response to a fungal elicitor, and initially down regulated and then up regulated in response to plant wounding (Sheng *et al.*, 1991). *DcAGP1* encodes an extracellular non-classical arabinogalactan-binding protein that may be involved in cell-cell interactions such as cell fate determination, cell division, cell expansion and cell death (Baldwin *et al.*, 2001). *CaPRP1* codes for a cell wall proline-rich glycoprotein which may play a role in the expansion processes of cells in rapidly elongating tissues or may be a structural cell wall matrix component (Mang *et al.*, 2004). The *NaPRP4* mRNA is detected specifically in the transmitting tissues of the styles, with transcript abundance correlating with elongation and maturation of the style, suggesting functions in maintenance of transmitting tissue integrity or providing nutrition or guidance for compatible pollen tubes (Chen *et al.*, 1993).

In addition to amino acid sequence homology, the cotton PRP shares several important structural similarities with the proteins encoded by *PvPRP1*, *DcAGP1*, *CaPRP1* and *NaPRP4*. They share a common domain structure, with PAC domains located within the C-terminal

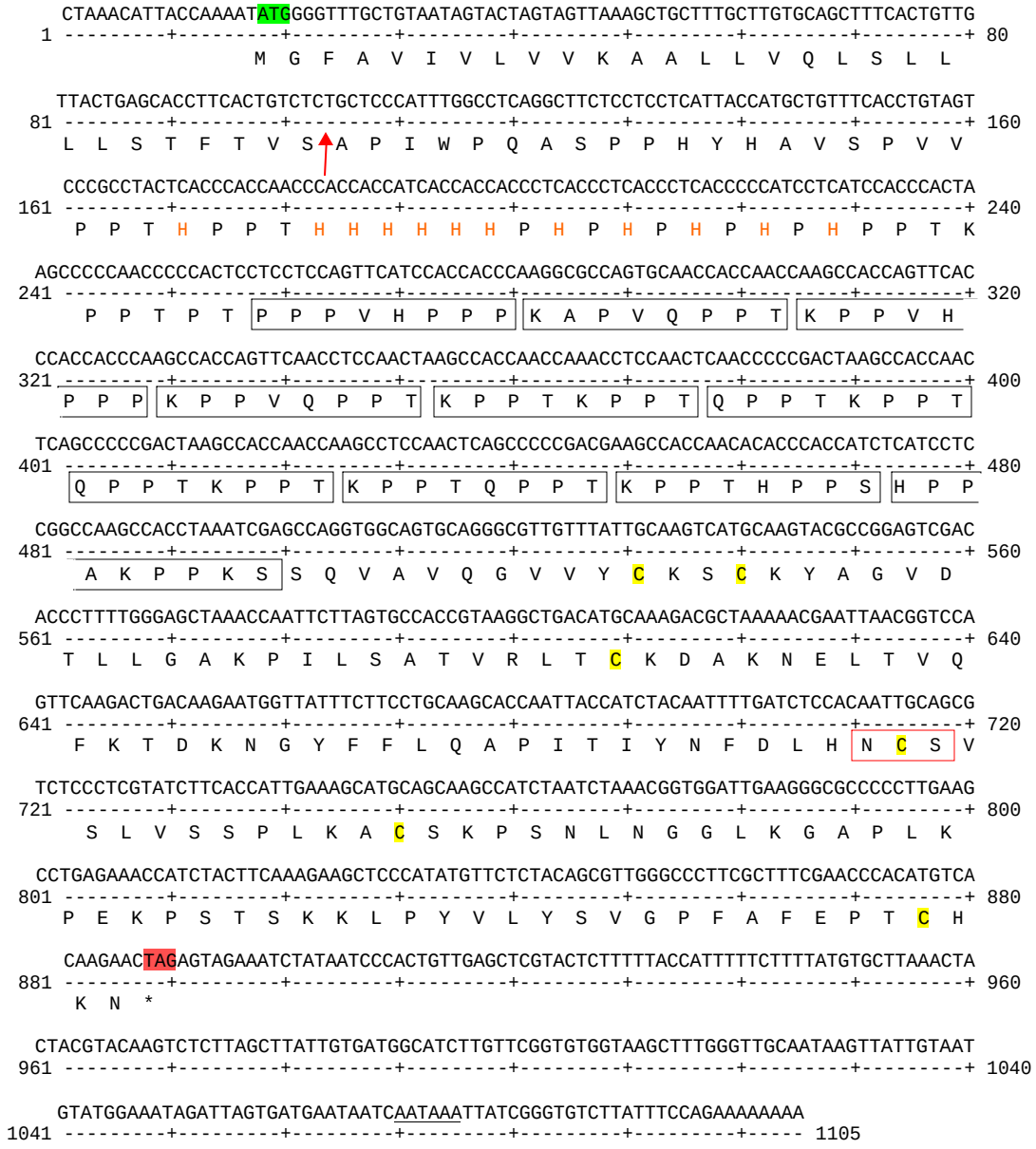


Figure 3.10: Nucleotide and conceptual amino acid sequences of proline rich protein. Nucleotide and conceptual amino acid sequences of the isolated PRP cDNA showing the putative translation start site (green shading) and the putative translation termination codon (red shading). The potential signal peptide cleavage site is indicated with a red arrow and the histidine residues that compose the histidine rich region are coloured orange. Proline rich repeats are boxed in black, a potential glycosylation site is boxed in red and the conserved cysteine residues are shaded in yellow. A potential polyadenylation signal site is underlined in black.

regions and proline rich N-terminal regions. Within the N-terminal regions, several common structural features are found including signal peptides, histidine-rich tracts and stretches of repeated amino acids rich in proline. The signal peptides are predicted to target the proteins either to the cell wall or outside the cell. The histidine-rich tracts may function as metal-binding sites or could form salt bridges with carboxyl groups on the galacturonosyl residues of pectin (Baldwin *et al.*, 2001). Histidine tracts are found within the PvPRP1 and DcAGP1 proteins, while CaPRP1 and NaPRP4 exhibit histidine-rich regions. Like the proline-rich repeats found in the N-terminal region of the identified cotton PRP, each protein contains N-terminal amino acid repeats of at least three residues in length involving proline. *PvPRP1* encodes a P-V-H-P-P-V-K-P-P-V repeat, *DcAGP1* encodes (V/T)-K-P-P repeats, *CaPRP1* codes for P-(S/T)-P-P-P repeats and *NaPRP4* encodes several K-P-P and P-T-K-P-P-T-Y-S-P-S-K-P-P repeats. Proline-rich repeats of this nature are typical of many cell wall-associated proteins but no function has yet been ascribed (Showalter, 2001).

A PAC domain is located in the C-terminal half of the cotton PRP, PvPRP1, DcAGP1, CaPRP1 and NaPRP4 (figure 3.10). The PAC domain is a domain of unknown function that is present in proline rich proteins and arabinogalactan proteins and is cysteine containing (Baldwin *et al.*, 2000). PAC domains usually contain six cysteine residues that are in conserved positions and are believed to form three disulphide bonds (Baldwin *et al.*, 2001). In addition to the conserved cysteines, potential glycosylation sites are also often present. The cotton PRP and related proteins contain similar PAC domains, with the six cysteine residues conserved. The only exception is PvPRP1, which contains only four of the six residues, suggesting that it may form only two of the three potential disulphide bonds (Baldwin *et al.*, 2001). While PAC domain function is unknown, several animal extracellular matrix proteins that have similar cysteine-rich domains are carbohydrate or glycoprotein binding (Fiete *et al.*, 1998). The conceptual protein encoded by the cotton PRP contains one potential N-glycosylation site (Asn-Cys-Ser) (figure 3.10) matching the consensus sequence of (Asn-X-Ser/Thr) (Wagh and Bahl, 1981), as does PvPRP1 (two sites: (Asn-Asn-Thr and Asn-Ala-Ser), DcAGP1 (two sites: Asn-Tyr-Thr and Asn-Asn-Thr), CaPRP1 (one site: Asn-Asn-Ser) and NaPRP4 (two sites: Asn-Asn-Thr and Asn-Asn-Thr). All of these potential N-glycosylation sites occur at or near conserved cysteine residues.

In addition to the structural similarities, the hydropathy profiles of the proteins are similar, with hydrophobic signal peptides followed by non-repetitive proline rich regions, repetitive proline rich regions that are hydrophilic, and C-terminal hydrophobic regions. All five proteins are strongly basic with isoelectric points ranging from 10.5 (PvPRP1) to 10.8

(CaPRP1), including the newly identified cotton PRP conceptual protein, which has an isoelectric point of 10.7.

The classification of the cotton PRP identified here is confused by the relative lack of similar, well-characterised plant PRPs and by the overlapping nomenclature applied to plant PRPs. The presence of a PAC domain and a potential N-glycosylation site suggests that the protein may bind carbohydrates, indicating that it may be a proline-rich glycoprotein. Sequence and structural similarity with other PRPs, which undergo extensive proline hydroxylation (to form hydroxyproline) and which are extensively O-glycosylated, suggests that the putative protein identified here may be a hydroxyproline-rich glycoprotein (HRGP) that may be involved in various aspects of plant growth and development. The HRGP family of proteins can be divided into PRPs, arabinogalactan proteins (AGP) and extensins. Baldwin *et al*. (2001) identified the protein encoded by *DcAGP1* as a non-classical AGP according to the classification of AGPs by Nothnagel (1997). Although PvPRP1 and NaPRP4 were not originally classified as non-classical AGPs, they have since been grouped into the same family (Baldwin *et al.*, 2001, Showalter, 2001). CaPRP1 was characterised as a proline-rich glycoprotein but it may also be a non-classical AGP, as it shows sequence and structural similarities with other AGP family members. Therefore, based on amino acid and structural similarity with previously characterised plant PRPs, the conceptual protein encoded by the cotton PRP identified here as appears to be a non-classical AGP within the broader family of glycosylated hydroxyproline-rich glycoproteins. Biochemical analysis could be used to confirm this classification. No definitive function has been ascribed to these proteins but it is believed that they may be involved in cell wall processes and various aspects of plant growth and development (Showalter, 2001). The suggestion that this protein interacts with the cell wall is supported by the presence of the transcript specifically in the boll wall tissue between 0 and 25 DPA (figure 3.7). The transcript is most highly abundant between 5 and 15 DPA, which coincides with the peak period of ball wall enlargement between 1 and 15 DPA (Wilkins and Jernstedt, 1999).

3.4.4: Chalcone synthase

Five of the cDNAs identified in the screen encode CHS. The largest of these clones is 1,466 bp in length, with 78 bp and 218 bp respectively of 3' UTR and 5' UTR and potential polyadenylation signals of TATTAA and TAATAA (figure 3.11). The 1,170 bp open reading frame conceptually encodes a protein of 389 amino acids with a mass of 43 kDa and an isoelectric point of 6.4, which is similar to the characteristics of other published CHS proteins (Claudot *et al.*, 1999). The cDNA does not contain a poly$(A)^+$ tail, but alignment with other

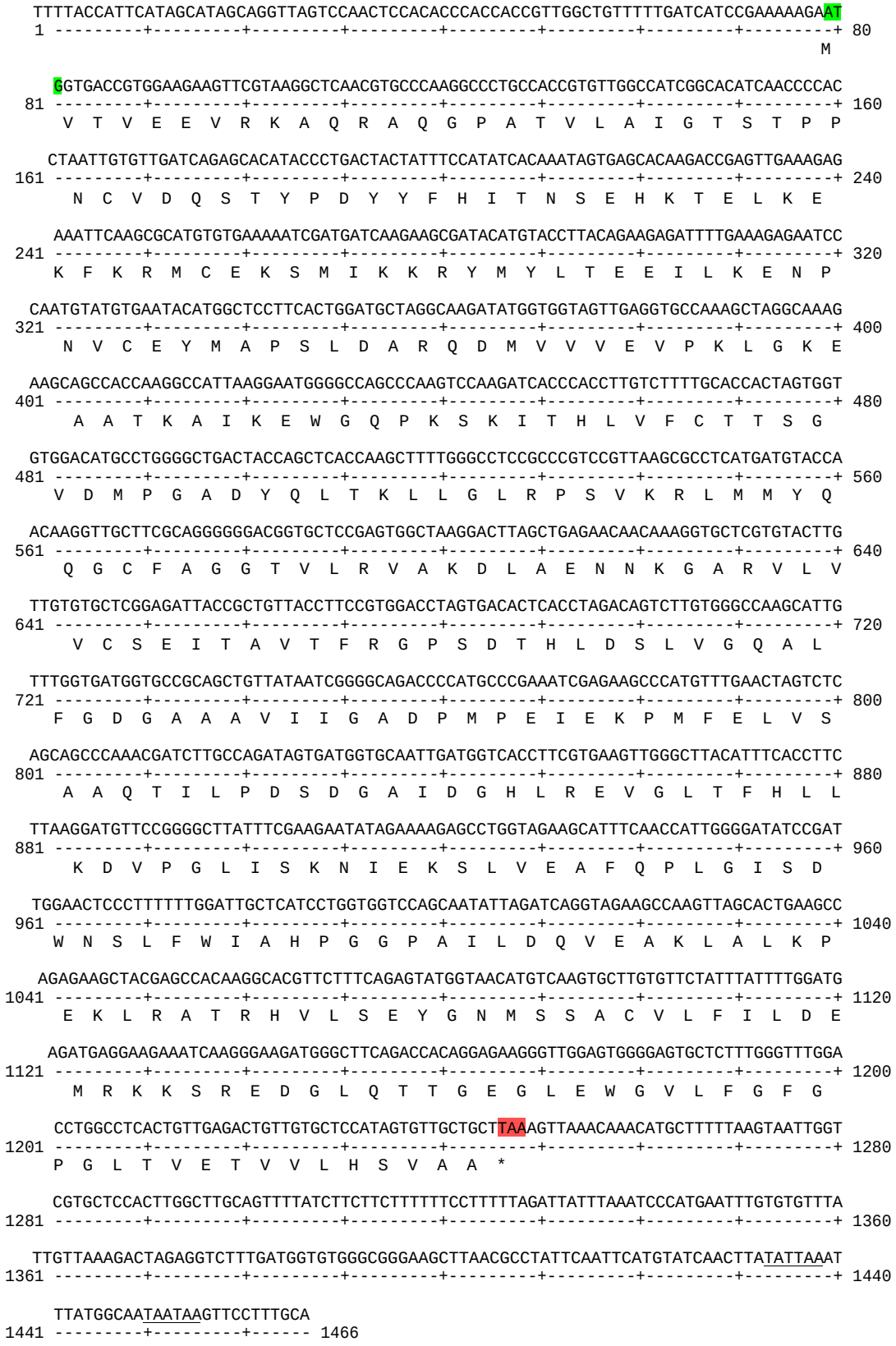


Figure 3.11: Nucleotide and conceptual amino acid sequences of chalcone synthase. Nucleotide and conceptual amino acid sequences of the isolated CHS cDNA, showing the putative translation start site (green shading) and the putative translation termination codon (red shading). Two potential polyadenylation signal sites are underlined in black.

CHS sequences indicates that a poly(A)$^+$ tail may be expected in close proximity to the end of the isolated cDNA. While no cotton CHS genes have been described previously, the conceptual protein shows more than 90% amino acid homology with CHS from many plant species including *Camellia sinensis* (Takeuchi *et al.*, 1994) *Juglans nigra* (Claudot *et al.*, 1999), *Rhododendron simsii* (GenBank accession number CAC88858) and *Vitis Vinifera* (Goto-Yamamoto *et al.*, 2002). Most plant species contain a small family of CHS genes consisting of one to three genes (Junghans *et al.*, 1993). Within a species, different CHS genes have been found to respond to different stimuli and are often expressed in different spatial or temporal patterns (Junghans *et al.*, 1993).

CHS catalyses the condensation of one molecule of 4-coumaroyl-CoA (from the phenylpropanoid pathway) with three molecules of malonyl-CoA to form naringenin chalcone in the first step of the flavonoid pathway (figure 3.12) (Yang *et al.*, 2002). It is a key enzyme in the flavonoid pathway, and is the most abundant and widely studied gene in the pathway (Richard *et al.*, 2000). The products of the flavonoid pathway include proteins that are involved in plant defence (proanthocyanidins), pigmentation (aurones, anthocyanins, flavonols and flavones), UV protection (anthocyanins and flavonols), pollen viability (flavonols), pollinator attractants (anthocyanins) and *Rhizobium* interaction (isoflavonoids) (reviewed by Dixon *et al.*, 2005). The expression of CHS is upregulated by UV light, pathogen attack, wounding and symbiotic rhizobia (Jahne *et al.*, 1993, Arioli *et al.*, 1994, Liu *et al.*, 1995, Reuber *et al.*, 1996).

There has been widespread interest in flavonoid genes in recent years because of their involvement in colour and taste traits in many plant species such as strawberry (Aharoni *et al.*, 2004) and grape (Bogs *et al.*, 2005). However, of more relevance to this study are outputs of the pathway that have roles in plant defence, such as the proanthocyanins (condensed tannins). Proanthocyanins have strong biological activities and can reduce insect nutrient absorption and cause midgut lesions (Peters and Constabel, 2002).

Given the roles for flavonoid pathway proteins in pest defence, it is not surprising to see a high level of CHS mRNA in the boll wall tissue between 0 and 20 DPA (figure 3.7) as up regulation of the pathway would lead to increase pest defence in the boll. The high level of CHS mRNA transcript in the petal tissue is also expected given mRNA expression data from other species and the involvement of the flavonoid pathway in petal colouration as reviewed by Koes *et al.* (2005).

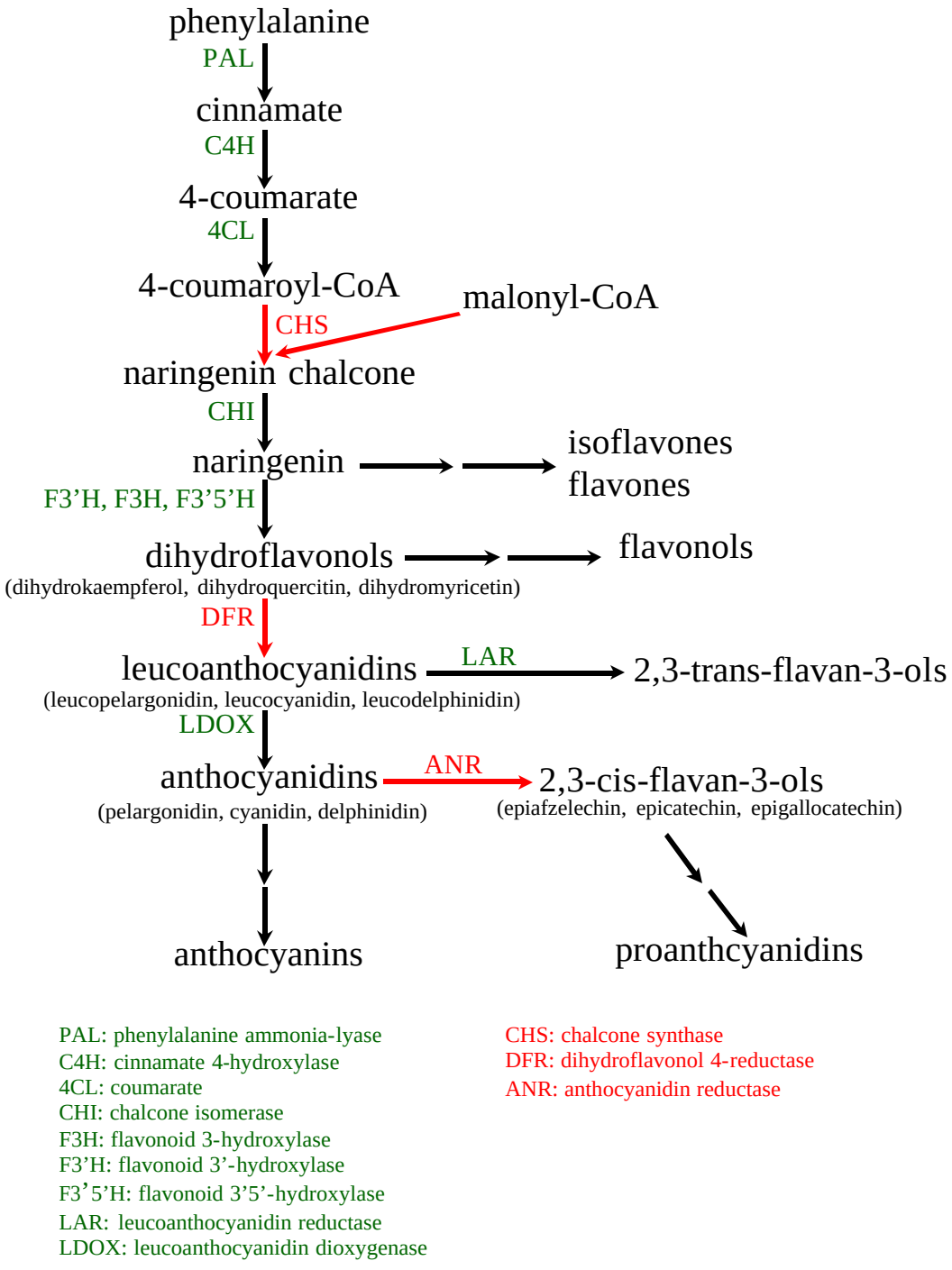


Figure 3.12: Simplified flavonoid pathway.
A simplified model of the flavonoid pathway of several plant species, with the proteins encoded by the mRNAs identified in this study shown in red. Enzyme names are abbreviated in the figure in either green or red, with full names listed below the pathway.

3.4.5: Dihydroflavonol 4-reductase

A single DFR cDNA was identified in the screen. The cDNA contains several possible start codons, with the sequence surrounding the potential ATG start site of translation (CATAATGGG) at nucleotide #82 in figure 3.13 being the closest to the consensus start site of translation (AAAAATGGC) for dicotyledon species (Joshi *et al.*, 1997). The clone is 1,276 bp long and contains an open reading frame of 1,023 bp, with an 81 bp 5' UTR and a 172 bp 3' UTR as well as a potential polyadenylation sequence of AATAAA (figure 3.13). The conceptual DFR protein is 340 amino acids in length, with a mass of 38 kDa and an isoelectric point of 5.9. It has between 70% and 80% homology with DFRs from several species including *Vitis vinifera* (GenBank accession number CAA72420), *Citrus sinensis* (Lo Piero *et al.*, 2006) and *Populus tremuloides* (Peters and Constabel, 2002).

DFR catalyses the reduction of dihydroflavonols to leucoanthocyanins. The reaction reduces dihydroflavonols such as dihydrokaempferol (DHK), dihydroquercetin (DHQ) and dihydromyricetin (DHM) to their corresponding leucoanthocyanidin products: leucopelargonidin, leucocyanidin and leucodelphinidin (figure 3.12). Most plant species contain between one and three copies of the DFR gene (Liu *et al.*, 2005, Shimada *et al.*, 2005), which show tissue-specific expression (Liu *et al.*, 2005) and are induced by developmental and environmental factors (Peters and Constabel, 2002).

Within the flavonoid pathway, the DFR enzyme functions downstream of CHS (figure 3.12) and the presence of cotton DFR mRNA in the boll wall between 0 and 15 DPA (figure 3.7) may indicate a role in production of proanthocyanins, which are involved in pest defence (Peters and Constabel, 2002).

3.4.6: Anthocyanidin reductase

A clone with sequence homology to ANR was identified once in the screen. The clone is 1,273 bp in length, with a 1,011 bp open reading frame. The transcript contains a 48 bp 3' UTR, a 214 bp 5' UTR and a potential polyadenylation sequence of TAATAT (figure 3.14). The open reading frame encodes a conceptual protein of 336 amino acids with a mass of 36 kDa and an isoelectric point of 5.3. The conceptual protein shows 99.4% homology with an ANR conceptual protein from *Gossypium arboreum* (Tanner *et al.*, 2003) and high homology with ANR conceptual proteins from several plant species including *Vitis vinifera,* (Fujita *et al.*, 2005), *Malus* x *domestica* (Takos *et al.*, 2006) and *Camellia sinensis* (GenBank accession number AAT68773).

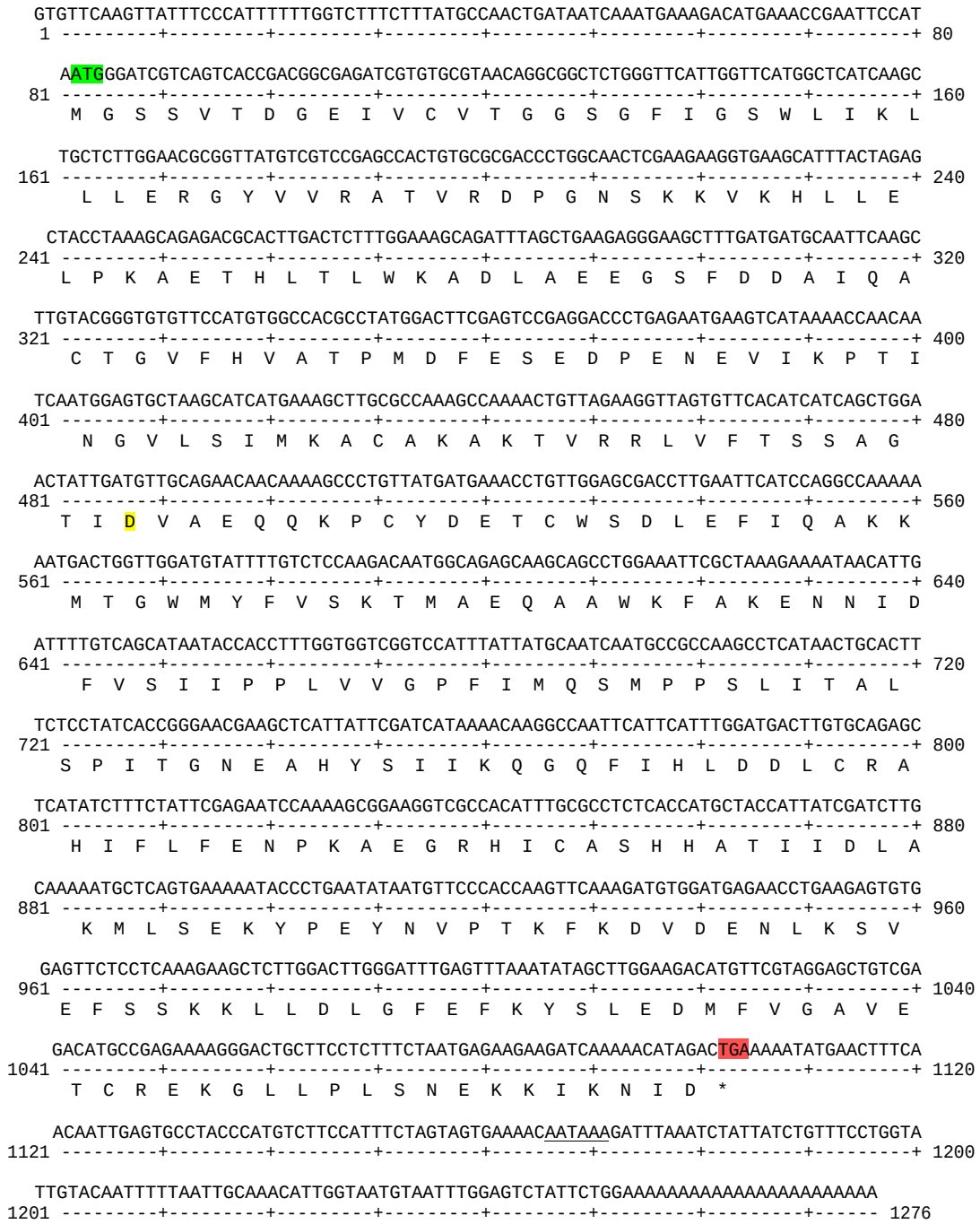


Figure 3.13: Nucleotide and conceptual amino acid sequences of dihydroflavonol 4-reductase. Nucleotide and conceptual amino acid sequences of the isolated DFR cDNA, showing the putative translation start site (green shading) and the putative translation termination codon (red shading). The aspartic acid residue that may be involved in substrate acceptance is shaded yellow. A potential polyadenylation signal site is underlined in black.

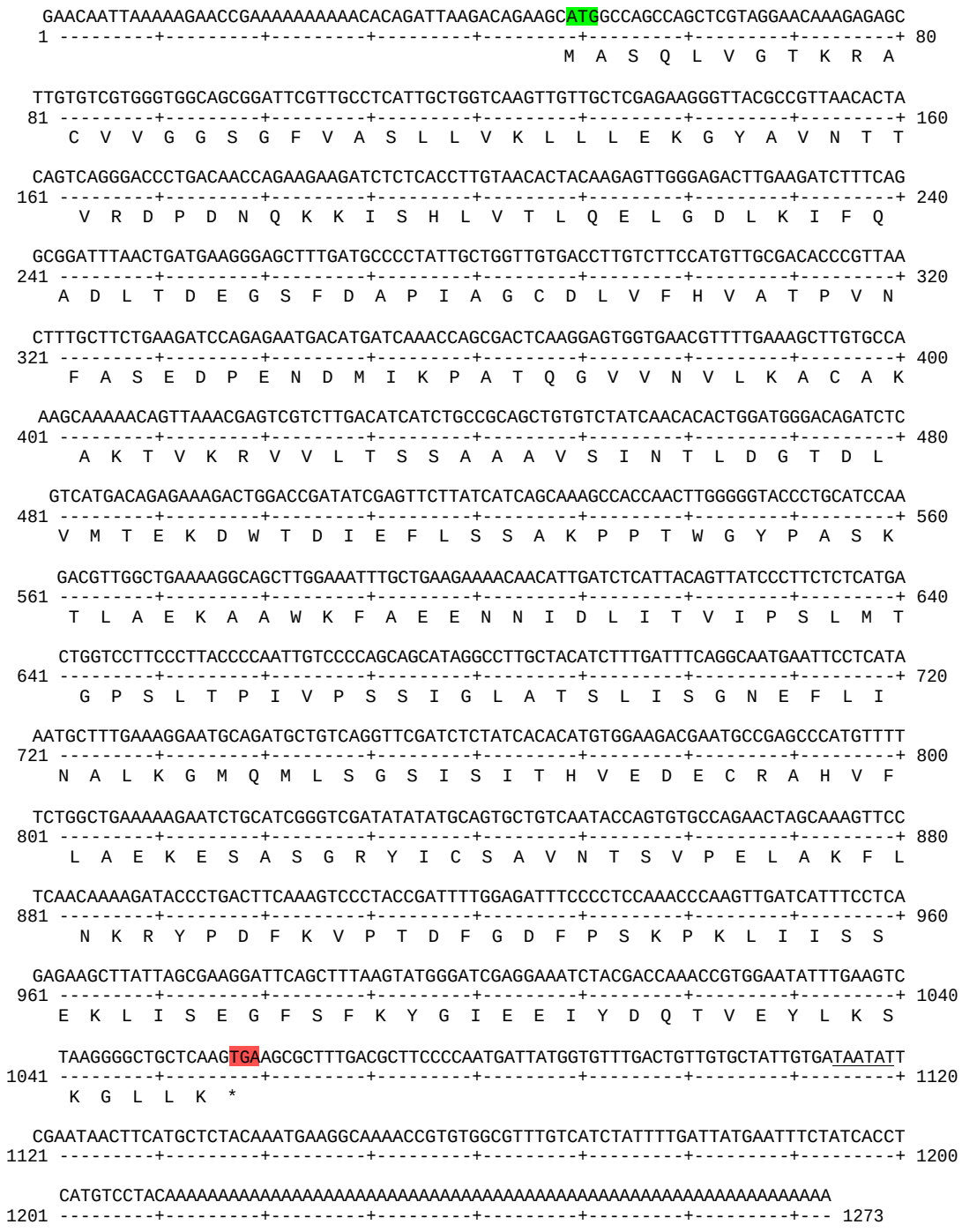


Figure 3.14: Nucleotide and conceptual amino acid sequences of anthocyanidin reductase. Nucleotide and conceptual amino acid sequences of the isolated ANR cDNA, showing the putative translation start site (green shading) and the putative translation termination codon (red shading). A potential polyadenylation signal site is underlined in black.

The ANR enzyme catalyses one of the two branch points in the flavonoid pathway that leads to proanthocyanidin production. ANR synthesises 2,3-cis-flavan-3-ols (epicatechin. epiafzelechin and epigallocatechin) from their corresponding anthocyanidins (cyanidin, pelargonidin and delphinidin) (figure 3.12) (Devic *et al.*, 1999). The other branch point in the flavonoid pathway that leads to proanthocyanidin production is catalysed by leucoanthocyanidin (LAR), which synthesises 2,3-trans-flavan-3-ols (Tanner *et al.*, 2003). 2,3-trans-flavan-3-ols, 2,3-cis-flavan-3-ols and leucoanthocyanidin are condensed together to form proanthocyanidin polymers (Xie *et al.*, 2003).

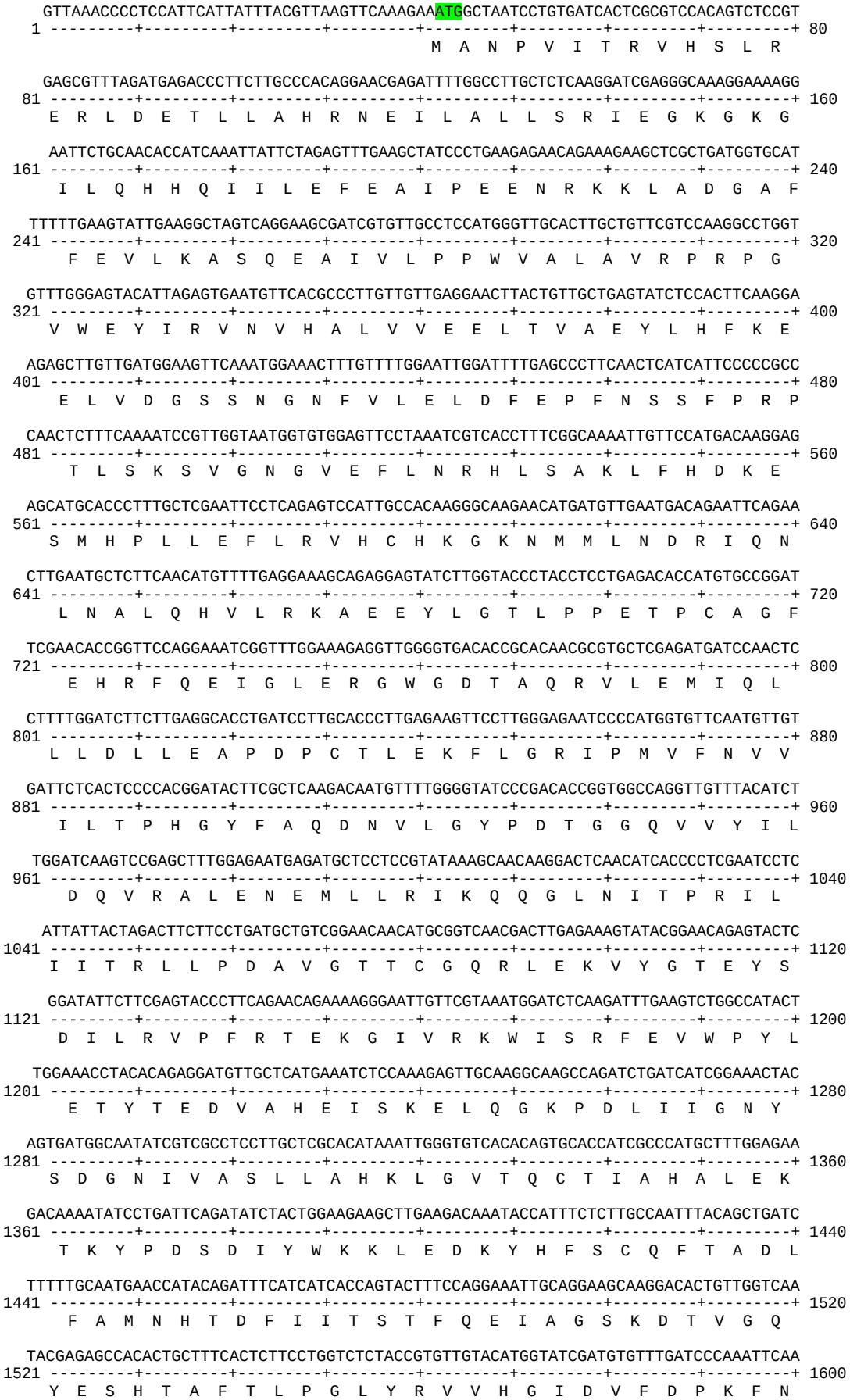
Within the flavonoid pathway, the ANR enzyme functions downstream of the CHS and DFR enzymes and represents a branch point in the pathway between the synthesis of anthocyanins and proanthocyanins (figure 3.12). Similar to the DFR transcript, the ANR mRNA was detected in the boll wall between 0 and 10 DPA (figure 3.7), suggesting a role in the production of proanthocyanins, which are known to have roles in pest defence (Peters and Constabel, 2002).

3.4.7: Sucrose synthase

Four of the clones identified from the differential screen encode SuSy. The longest of these is 2,646 bp in length, contains a potential open reading frame of 2,418 bp, a 41 bp 5' UTR and a 187 bp 3' UTR. Several potential polyadenylation signals are located within the 3' UTR (figure 3.15). By comparison with another SuSy clone with a longer 3' cDNA end, the last adenine of this clone appears to be the beginning of the poly(A)$^+$ tail. The putative protein encoded by this open reading frame is 805 amino acids long with a mass of 93 kDa and an isoelectric point of 6.6.

The clone is identical to a cotton SuSy cDNA identified by Wu *et al*. (2006) and 99% similar to another cotton SuSy cDNA (GenBank accession number AAD28641). The SuSy enzyme catalyses a reversible reaction, but preferentially converts sucrose and UDP to fructose and UDP-glucose (Geigenberger and Stitt, 1993). SuSy has been extensively studied in many plant species and has been shown to play a major role in energy metabolism and mobilisation of sucrose into various pathways that are important for the metabolic, structural and storage functions of the plant cell (Baud *et al.*, 2004).

SuSy has been well characterised in cotton because of its role in fibre growth and development. During the rapid elongation and expansion of cotton fibres, a cellulose-rich


GTTAAACCCCTCCATTCATTATTTACGTTAAGTTCAAAGAAATGGCTAATCCTGTGATCACTCGCGTCCACAGTCTCCGT
1 ---------+---------+---------+---------+---------+---------+---------+---------+ 80
M A N P V I T R V H S L R
GAGCGTTTAGATGAGACCCTTCTTGCCCACAGGAACGAGATTTTGGCCTTGCTCTCAAGGATCGAGGGCAAAGGAAAAGG
81 ---------+---------+---------+---------+---------+---------+---------+---------+ 160
E R L D E T L L A H R N E I L A L L S R I E G K G K G
AATTCTGCAACACCATCAAATTATTCTAGAGTTTGAAGCTATCCCTGAAGAGAACAGAAAGAAGCTCGCTGATGGTGCAT
161 ---------+---------+---------+---------+---------+---------+---------+---------+ 240
I L Q H H Q I I L E F E A I P E E N R K K L A D G A F
TTTTTGAAGTATTGAAGGCTAGTCAGGAAGCGATCGTGTTGCCTCCATGGGTTGCACTTGCTGTTCGTCCAAGGCCTGGT
241 ---------+---------+---------+---------+---------+---------+---------+---------+ 320
F E V L K A S Q E A I V L P P W V A L A V R P R P G
GTTTGGGAGTACATTAGAGTGAATGTTCACGCCCTTGTTGTTGAGGAACTTACTGTTGCTGAGTATCTCCACTTCAAGGA
321 ---------+---------+---------+---------+---------+---------+---------+---------+ 400
V W E Y I R V N V H A L V V E E L T V A E Y L H F K E
AGAGCTTGTTGATGGAAGTTCAAATGGAAACTTTGTTTTGGAATTGGATTTTGAGCCCTTCAACTCATCATTCCCCCGCC
401 ---------+---------+---------+---------+---------+---------+---------+---------+ 480
E L V D G S S N G N F V L E L D F E P F N S S F P R P
CAACTCTTTCAAAATCCGTTGGTAATGGTGTGGAGTTCCTAAATCGTCACCTTTCGGCAAAATTGTTCCATGACAAGGAG
481 ---------+---------+---------+---------+---------+---------+---------+---------+ 560
T L S K S V G N G V E F L N R H L S A K L F H D K E
AGCATGCACCCTTTGCTCGAATTCCTCAGAGTCCATTGCCACAAGGGCAAGAACATGATGTTGAATGACAGAATTCAGAA
561 ---------+---------+---------+---------+---------+---------+---------+---------+ 640
S M H P L L E F L R V H C H K G K N M M L N D R I Q N
CTTGAATGCTCTTCAACATGTTTTGAGGAAAGCAGAGGAGTATCTTGGTACCCTACCTCCTGAGACACCATGTGCCGGAT
641 ---------+---------+---------+---------+---------+---------+---------+---------+ 720
L N A L Q H V L R K A E E Y L G T L P P E T P C A G F
TCGAACACCGGTTCCAGGAAATCGGTTTGGAAAGAGGTTGGGGTGACACCGCACAACGCGTGCTCGAGATGATCCAACTC
721 ---------+---------+---------+---------+---------+---------+---------+---------+ 800
E H R F Q E I G L E R G W G D T A Q R V L E M I Q L
CTTTTGGATCTTCTTGAGGCACCTGATCCTTGCACCCTTGAGAAGTTCCTTGGGAGAATCCCCATGGTGTTCAATGTTGT
801 ---------+---------+---------+---------+---------+---------+---------+---------+ 880
L L D L L E A P D P C T L E K F L G R I P M V F N V V
GATTCTCACTCCCCACGGATACTTCGCTCAAGACAATGTTTTGGGGTATCCCGACACCGGTGGCCAGGTTGTTTACATCT
881 ---------+---------+---------+---------+---------+---------+---------+---------+ 960
I L T P H G Y F A Q D N V L G Y P D T G G Q V V Y I L
TGGATCAAGTCCGAGCTTTGGAGAATGAGATGCTCCTCCGTATAAAGCAACAAGGACTCAACATCACCCCTCGAATCCTC
961 ---------+---------+---------+---------+---------+---------+---------+---------+ 1040
D Q V R A L E N E M L L R I K Q Q G L N I T P R I L
ATTATTACTAGACTTCTTCCTGATGCTGTCGGAACAACATGCGGTCAACGACTTGAGAAAGTATACGGAACAGAGTACTC
1041 ---------+---------+---------+---------+---------+---------+---------+---------+ 1120
I I T R L L P D A V G T T C G Q R L E K V Y G T E Y S
GGATATTCTTCGAGTACCCTTCAGAACAGAAAAGGGAATTGTTCGTAAATGGATCTCAAGATTTGAAGTCTGGCCATACT
1121 ---------+---------+---------+---------+---------+---------+---------+---------+ 1200
D I L R V P F R T E K G I V R K W I S R F E V W P Y L
TGGAAACCTACACAGAGGATGTTGCTCATGAAATCTCCAAAGAGTTGCAAGGCAAGCCAGATCTGATCATCGGAAACTAC
1201 ---------+---------+---------+---------+---------+---------+---------+---------+ 1280
E T Y T E D V A H E I S K E L Q G K P D L I I G N Y
AGTGATGGCAATATCGTCGCCTCCTTGCTCGCACATAAATTGGGTGTCACACAGTGCACCATCGCCCATGCTTTGGAGAA
1281 ---------+---------+---------+---------+---------+---------+---------+---------+ 1360
S D G N I V A S L L A H K L G V T Q C T I A H A L E K
GACAAAATATCCTGATTCAGATATCTACTGGAAGAAGCTTGAAGACAAATACCATTTCTCTTGCCAATTTACAGCTGATC
1361 ---------+---------+---------+---------+---------+---------+---------+---------+ 1440
T K Y P D S D I Y W K K L E D K Y H F S C Q F T A D L
TTTTTGCAATGAACCATACAGATTTCATCATCACCAGTACTTTCCAGGAAATTGCAGGAAGCAAGGACACTGTTGGTCAA
1441 ---------+---------+---------+---------+---------+---------+---------+---------+ 1520
F A M N H T D F I I T S T F Q E I A G S K D T V G Q
TACGAGAGCCACACTGCTTTCACTCTTCCTGGTCTCTACCGTGTTGTACATGGTATCGATGTGTTTGATCCCAAATTCAA
1521 ---------+---------+---------+---------+---------+---------+---------+---------+ 1600
Y E S H T A F T L P G L Y R V V H G I D V F D P K F N

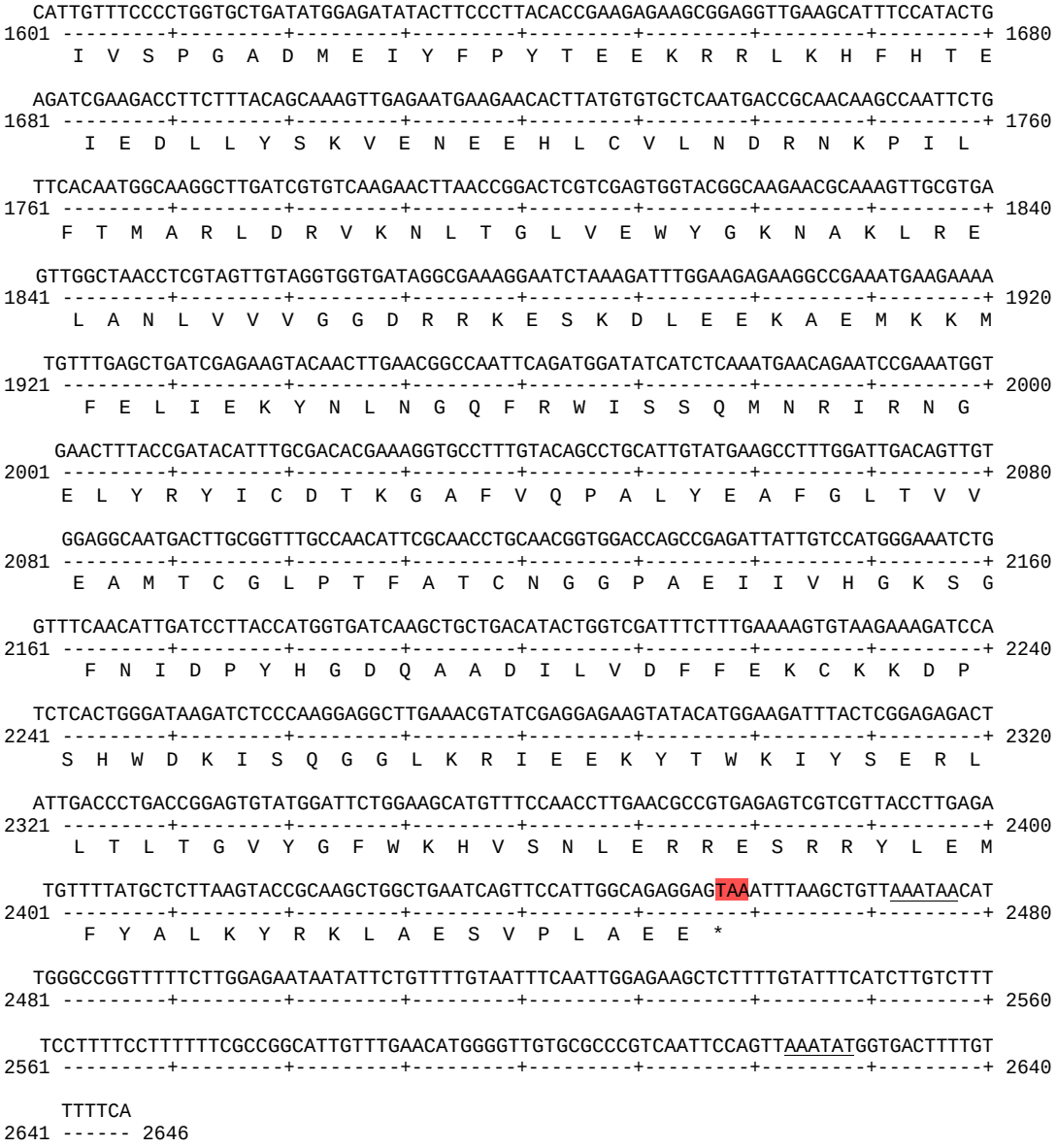


Figure 3.15: Nucleotide and conceptual amino acid sequences of sucrose synthase. Nucleotide and conceptual amino acid sequences of the isolated SuSy cDNA, showing the putative translation start site (green shading) and the putative translation termination codon (red shading). Two potential polyadenylation signal sites are underlined in black.

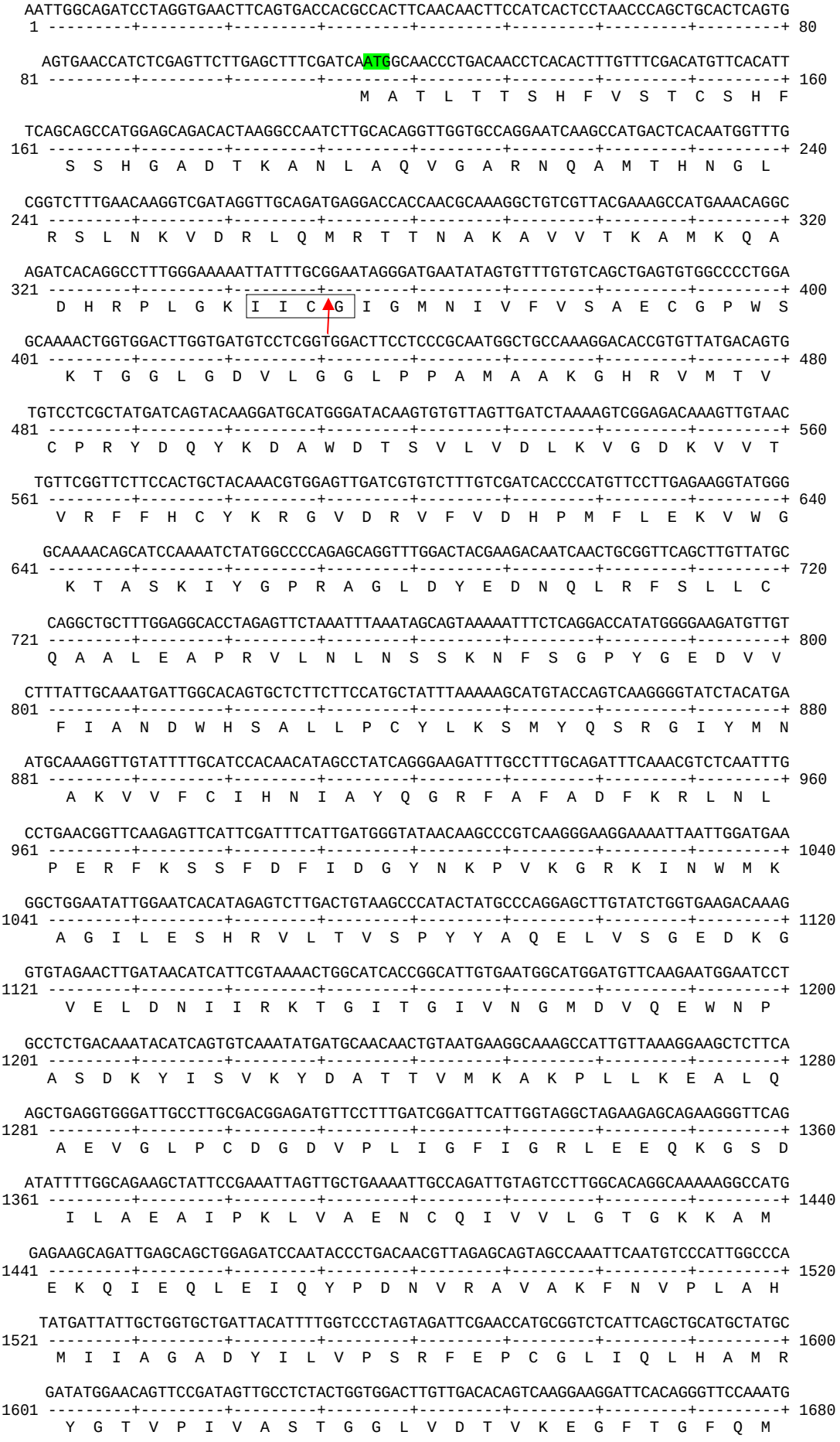
secondary cell wall is deposited (Schubert *et al.*, 1973). To support this cell wall deposition, the fibres act as a sink for sucrose, which is then utilised for fibre expansion and cellulose synthesis (Haigler *et al.*, 2001, Ruan *et al.*, 2001). In fibres, SuSy is the major enzyme that degrades imported sucrose (Ruan and Chourey, 1998) to generate UDP-Glucose, which is then channelled into cellulose synthesis (Haigler *et al.*, 2001). In addition to the role of SuSy in cellulose synthesis, fructose and UDP-glucose produced by SuSy can also be used in energy generation and in the biosynthesis of other products such as starch (Ruan *et al.*, 2005). SuSy activity correlates with the sink strength of storage organs, providing substrates for starch synthesis in potato tubers (Ross *et al.*, 1994), tomato fruits (D'Aoust *et al.*, 1999), maize kernels (Baud *et al.*, 2004) and pea embryos (Dejardin *et al.*, 1997).

Northern analysis showed that the SuSy mRNA is present in most tissues including 0 to 25 DPA boll wall, leaf, fibre, white film, bract, calyx, petal, staminal column, root and stem, with highest abundance in the early boll wall timepoints (figure 3.7). The high abundance of SuSy mRNA in the boll wall corresponds with the timing of growth and secondary cell wall synthesis of the cotton fibres. SuSy may function in the boll wall to catalyse the breakdown of sucrose to fructose and thereby supply the fibres with UDP-glucose for cellulose synthesis. Furthermore, the breakdown of sucrose in the boll may increase the osmotic potential for sucrose import into the boll.

3.4.8: Granule-bound starch synthase

Three clones that were isolated in the screen encode proteins of the GBSS class. The longest of these clones is 2,258 bp in length with a potential open reading frame of 1,830 bp. The transcript contains a 114 bp 5' UTR, a 314 bp 3' UTR and a polyadenylation signal 19 bp before the poly(A)$^+$ tail (figure 3.16). The conceptual protein shows between 75% and 77% homology with GBSS proteins from several plant species including *Manihot esculenta* (Salehuzzaman *et al.*, 1993), *Antirrhinum majus* (GenBank accession number CAA06958) and *Solanum tuberosum (Vanderleij et al.,* 1991*)* and contains a signal peptide as indicated by the conserved cleavage site motif (IICG) (Salehuzzaman *et al.*, 1993) (figure 3.16). The transit peptide is 77 amino acids long with cleavage expected between the cysteine and glycine residues, with the resultant mature protein targeted to chloroplasts or amyloplasts. The cleaved conceptual protein is 530 amino acids in length with a mass of 58 kDa and an isoelectric point of 7.1.

Starch is an insoluble polymer composed of glucose residues and is a major energy storage molecule in plants (Geigenberger, 2003, Tenorio *et al.*, 2003, Tetlow *et al.*, 2004). It


AATTGGCAGATCCTAGGTGAACTTCAGTGACCACGCCACTTCAACAACTTCCATCACTCCTAACCCAGCTGCACTCAGTG
1 --------+---------+---------+---------+---------+---------+---------+---------+ 80
AGTGAACCATCTCGAGTTCTTGAGCTTTCGATCAATGGCAACCCTGACAACCTCACACTTTGTTTCGACATGTTCACATT
81 --------+---------+---------+---------+---------+---------+---------+---------+ 160
M A T L T T S H F V S T C S H F
TCAGCAGCCATGGAGCAGACACTAAGGCCAATCTTGCACAGGTTGGTGCCAGGAATCAAGCCATGACTCACAATGGTTTG
161 --------+---------+---------+---------+---------+---------+---------+---------+ 240
S S H G A D T K A N L A Q V G A R N Q A M T H N G L
CGGTCTTTGAACAAGGTCGATAGGTTGCAGATGAGGACCACCAACGCAAAGGCTGTCGTTACGAAAGCCATGAAACAGGC
241 --------+---------+---------+---------+---------+---------+---------+---------+ 320
R S L N K V D R L Q M R T T N A K A V V T K A M K Q A
AGATCACAGGCCTTTGGGAAAAATTATTTGCGGAATAGGGATGAATATAGTGTTTGTGTCAGCTGAGTGTGGCCCCTGGA
321 --------+---------+---------+---------+---------+---------+---------+---------+ 400
D H R P L G K I I C G I G M N I V F V S A E C G P W S
GCAAAACTGGTGGACTTGGTGATGTCCTCGGTGGACTTCCTCCCGCAATGGCTGCCAAAGGACACCGTGTTATGACAGTG
401 --------+---------+---------+---------+---------+---------+---------+---------+ 480
K T G G L G D V L G G L P P A M A A K G H R V M T V
TGTCCTCGCTATGATCAGTACAAGGATGCATGGGATACAAGTGTGTTAGTTGATCTAAAAGTCGGAGACAAAGTTGTAAC
481 --------+---------+---------+---------+---------+---------+---------+---------+ 560
C P R Y D Q Y K D A W D T S V L V D L K V G D K V V T
TGTTCGGTTCTTCCACTGCTACAAACGTGGAGTTGATCGTGTCTTTGTCGATCACCCCATGTTCCTTGAGAAGGTATGGG
561 --------+---------+---------+---------+---------+---------+---------+---------+ 640
V R F F H C Y K R G V D R V F V D H P M F L E K V W G
GCAAAACAGCATCCAAAATCTATGGCCCCAGAGCAGGTTTGGACTACGAAGACAATCAACTGCGGTTCAGCTTGTTATGC
641 --------+---------+---------+---------+---------+---------+---------+---------+ 720
K T A S K I Y G P R A G L D Y E D N Q L R F S L L C
CAGGCTGCTTTGGAGGCACCTAGAGTTCTAAATTTAAATAGCAGTAAAAATTTCTCAGGACCATATGGGGAAGATGTTGT
721 --------+---------+---------+---------+---------+---------+---------+---------+ 800
Q A A L E A P R V L N L N S S K N F S G P Y G E D V V
CTTTATTGCAAATGATTGGCACAGTGCTCTTCTTCCATGCTATTTAAAAAGCATGTACCAGTCAAGGGGTATCTACATGA
801 --------+---------+---------+---------+---------+---------+---------+---------+ 880
F I A N D W H S A L L P C Y L K S M Y Q S R G I Y M N
ATGCAAAGGTTGTATTTTGCATCCACAACATAGCCTATCAGGGAAGATTTGCCTTTGCAGATTTCAAACGTCTCAATTTG
881 --------+---------+---------+---------+---------+---------+---------+---------+ 960
A K V V F C I H N I A Y Q G R F A F A D F K R L N L
CCTGAACGGTTCAAGAGTTCATTCGATTTCATTGATGGGTATAACAAGCCCGTCAAGGGAAGGAAAATTAATTGGATGAA
961 --------+---------+---------+---------+---------+---------+---------+---------+ 1040
P E R F K S S F D F I D G Y N K P V K G R K I N W M K
GGCTGGAATATTGGAATCACATAGAGTCTTGACTGTAAGCCCATACTATGCCCAGGAGCTTGTATCTGGTGAAGACAAAG
1041 --------+---------+---------+---------+---------+---------+---------+---------+ 1120
A G I L E S H R V L T V S P Y Y A Q E L V S G E D K G
GTGTAGAACTTGATAACATCATTCGTAAAACTGGCATCACCGGCATTGTGAATGGCATGGATGTTCAAGAATGGAATCCT
1121 --------+---------+---------+---------+---------+---------+---------+---------+ 1200
V E L D N I I R K T G I T G I V N G M D V Q E W N P
GCCTCTGACAAATACATCAGTGTCAAATATGATGCAACAACTGTAATGAAGGCAAAGCCATTGTTAAAGGAAGCTCTTCA
1201 --------+---------+---------+---------+---------+---------+---------+---------+ 1280
A S D K Y I S V K Y D A T T V M K A K P L L K E A L Q
AGCTGAGGTGGGATTGCCTTGCGACGGAGATGTTCCTTTGATCGGATTCATTGGTAGGCTAGAAGAGCAGAAGGGTTCAG
1281 --------+---------+---------+---------+---------+---------+---------+---------+ 1360
A E V G L P C D G D V P L I G F I G R L E E Q K G S D
ATATTTTGGCAGAAGCTATTCCGAAATTAGTTGCTGAAAATTGCCAGATTGTAGTCCTTGGCACAGGCAAAAAGGCCATG
1361 --------+---------+---------+---------+---------+---------+---------+---------+ 1440
I L A E A I P K L V A E N C Q I V V L G T G K K A M
GAGAAGCAGATTGAGCAGCTGGAGATCCAATACCCTGACAACGTTAGAGCAGTAGCCAAATTCAATGTCCCATTGGCCCA
1441 --------+---------+---------+---------+---------+---------+---------+---------+ 1520
E K Q I E Q L E I Q Y P D N V R A V A K F N V P L A H
TATGATTATTGCTGGTGCTGATTACATTTTGGTCCCTAGTAGATTCGAACCATGCGGTCTCATTCAGCTGCATGCTATGC
1521 --------+---------+---------+---------+---------+---------+---------+---------+ 1600
M I I A G A D Y I L V P S R F E P C G L I Q L H A M R
GATATGGAACAGTTCCGATAGTTGCCTCTACTGGTGGACTTGTTGACACAGTCAAGGAAGGATTCACAGGGTTCCAAATG
1601 --------+---------+---------+---------+---------+---------+---------+---------+ 1680
Y G T V P I V A S T G G L V D T V K E G F T G F Q M

```
     GGAGCTTTCAATGTTGAATGTGATGAAGTGGATCCAAGTGATGTGATTAAGGTGGTAAAAACTGTCAAGAGAGCTCTTGC
1681 ---------+---------+---------+---------+---------+---------+---------+---------+ 1760
     G  A  F  N  V  E  C  D  E  V  D  P  S  D  V  I  K  V  V  K  T  V  K  R  A  L  A

   AACATATGGGACTCAAGCACTGAAAGAAATGATCCAGAATTGTATGGCACAAGATTTTTCATGGAAGGGACCATCAAGGT
1761 ---------+---------+---------+---------+---------+---------+---------+---------+ 1840
      T  Y  G  T  Q  A  L  K  E  M  I  Q  N  C  M  A  Q  D  F  S  W  K  G  P  S  R  L

   TGTGGGAGAAGATGTTGTTGAGCCTCGGGGTGGCTGGCAGTGAACCTGGCATTGAAGGAGAGGAGGTTGCTCCTCTTGCC
1841 ---------+---------+---------+---------+---------+---------+---------+---------+ 1920
       W  E  K  M  L  L  S  L  G  V  A  G  S  E  P  G  I  E  G  E  E  V  A  P  L  A

     AAGGAAAATGTTGCCACACCCTGAGCATAATCAGAAGGCAACTATGGATAAAGCTTAGTAACCTGTTTTCTTAGACAACC
1921 ---------+---------+---------+---------+---------+---------+---------+---------+ 2000
     K  E  N  V  A  T  P  *

   AAGTCCAAGTACCTGTCTACTCTCGATAGCCATTGGCTTGACCGGCGGAAAACAAACTGACCTGCCATACGTGAAGTCAG
2001 ---------+---------+---------+---------+---------+---------+---------+---------+ 2080

    TAGTGCTATTTTTGAGGGTGAAAAGTCAGGAGATAGTATGATATATATCCATAGTTATAGCAGCTAATGGGTGTTAGTAA
2081 ---------+---------+---------+---------+---------+---------+---------+---------+ 2160

   TGTTGTCAAGAACTTAAAAAGAGAATCAGTTTCTCTGCATGGTTTTATGAGTAGTGAGCAATGTGAAATAAGATAAAATG
2161 ---------+---------+---------+---------+---------+---------+---------+---------+ 2240

     TTCAACTTCTAAAAAAAA
2241 ---------+-------- 2258
```

Figure 3.16: Nucleotide and conceptual amino acid sequences of granule-bound starch synthase. Nucleotide and conceptual amino acid sequences of the isolated GBSS cDNA, showing the putative translation start site (green shading) and the putative translation termination codon (red shading). The potential signal peptide cleavage site motif is indicated with a black box, with the cleavage site designated by a red arrow. A potential polyadenylation signal site is underlined in black.

is synthesised inside plastids and forms large complexes known as starch granules, which are composed of amylose and amylopectin. Amylose is comprised of largely unbranched α-1,4 glucan chains, while amylopectin is more highly branched and accounts for approximately 75% of the granule mass. The granule structure is produced by the formation of α-1,6 linkages between adjoining amylose and amylopectin chains (Smith *et al.*, 1997), which then pack together in organised structures (reviewed by Buleon *et al.*, 1998; Thompson, 2000). Starch synthase enzymes extend starch polymers by transferring the glycosyl group of ADP-glucose to the non-reducing ends of pre-existing starch molecules by establishing new α-1,4 bonds.

There are two classes of starch synthase enzymes: soluble starch synthase (SSS) and GBSS. Biochemical and genetic studies have shown that GBSS is predominantly responsible for the synthesis of amylose, while SSS is predominantly responsible for the synthesis of amylopectin (Ball *et al.*, 1998). GBSS is generally located within the granule matrix while SSS is located within plastid stroma (Hirose and Terao, 2004).

The mRNA of the GBSS identified here was detected predominantly in the boll wall tissue (0 to 40 DPA) with minor levels also observed in leaf, fibre, bract and calyx tissues (figure 3.7). Plants commonly use starch as a storage molecule for excess carbohydrate. For example, carbohydrates produced by photosynthesis in the leaves are exported to other, less photosynthetically active plant structures (Roitsch, 1999), including the boll (Bondada *et al.*, 1994), where the carbohydrates are converted into starch. The high abundance of GBSS mRNA in the boll wall suggests that the cotton boll is synthesising and storing large quantities of starch, which may be associated with the presence of starch-rich plastids within the boll wall.

3.5: Relative abundance of the boll wall-specific transcripts

The method of differential screening is biased towards the identification of highly abundant cDNAs due to the nature of the cDNA probes. Clones within the library that correspond to cDNAs that are of low abundance within the probe population will not be labelled with sufficient radioactive molecules to be detectable (Sambrook *et al.*, 1989). mRNA transcripts of greater than 0.05% abundance in the target tissue (boll wall) are usually identified by this type of screen (Sambrook *et al.*, 1989).

A representative portion of the library was screened with each of the six isolated cDNAs to determine the number of similar clones within the library and hence the approximate abundance of each of the cDNAs within the library. The relative abundance of

the transcripts provides an approximation of the strength of each promoter. The abundance of each of the six cDNAs within the library is shown in table 3.1.

cDNA identity	Clones in class	Clones screened	Positive clones	Abundance
Chalcone synthase	4	6000	18	0.30%
Sucrose synthase	5	6000	15	0.25%
Proline rich protein	6	6000	10	0.17%
Starch synthase	3	6000	9	0.15%
Dihydroflavonol 4-reductase	1	9000	7	0.08%
Anthocyanidin reductase	1	9000	6	0.07%

Table 3.1: Relative abundance of the six boll wall-specific transcripts.
As an approximation of promoter strength, the number of clones of each class within a representative portion of the 5 DPA boll wall cDNA library was determined by hybridising plaque lifts of the library with each of the cDNAs.

As expected, each of the six cDNAs was present within the library at greater than 0.05% abundance, ranging from 0.07% to 0.30%. Cumulatively, the transcripts account for a significant proportion of the library (1.01%). These estimations assume that each cDNA is equally cloneable, each clone contains a single cDNA and that the hybridisation is highly specific (i.e. there is no cross hybridisation to similar transcripts). However, several colonies isolated amongst the initial 30 clones contained more than one cDNA, meaning that the estimations are probably exaggerated.

The abundance of each of the cDNAs in the library correlates well with the level of mRNA transcript detected in the 5 DPA boll wall timepoint by Northern analysis (figure 3.7). CHS and SuSy transcripts are the most highly abundant, accounting for the largest proportions of the cDNAs within the library. The PRP and GBSS transcripts are less abundant with moderate mRNA levels and cDNA abundance within the library. DFR and ANR transcripts have the lowest transcript abundance in the boll wall at 5 DPA according to the Northern analysis and account for the lowest proportion of cDNAs in the library. The abundance of the cDNAs within the library also correlates well with the number of clones of each class that were initially identified in the primary screen (table 3.1). The clones that were isolated multiple times from the initial screen have the highest mRNA transcript abundance as indicated by Northern analysis and also account for the largest proportions of the cDNA library.

3.6: Summary and conclusions

From the differential screen of a 5 DPA boll wall cDNA library, 30 cDNA clones were identified that were much more abundant in boll wall than leaf tissue. Six major classes of cDNA were identified within these initial clones that correspond to mRNA transcripts that are highly specific to the boll wall, namely PRP, GBSS, DFR, ANR, CHS and SuSy. The cDNAs were isolated and characterised, with five of the cDNAs (PRP, GBSS, DFR, ANR and CHS) corresponding to cotton genes that have not been previously characterised.

Northern analysis was utilised to determine the abundance of each of the transcripts within various tissues. Each of the transcripts is most highly abundant in the boll wall tissue, with lower levels detected in several other tissues. The PRP, GBSS, CHS and SuSy transcripts are present at relatively high levels in the boll wall, while the DFR and ANR transcripts are less abundant. The PRP transcript has the most restricted expression with no transcript detected in non-boll wall tissues.

This chapter describes the identification of abundant mRNA transcripts that are highly specific to the boll wall tissue of cotton. In the following chapter, the identification and isolation of the promoters that correspond to several of the transcripts will be described.

Chapter 4: Identification and characterisation of cotton boll wall-specific promoters

4.1: Introduction

The aim of this project is to identify and characterise cotton promoters that drive boll wall-specific gene expression. In the previous chapter, a 5 DPA boll wall cDNA library was differentially screened with 5 DPA boll wall and leaf cDNA probes to identify transcripts that are significantly more abundant in, or specific to, boll wall tissue. Six distinct classes of cDNA were isolated and partially characterised. Three abundant mRNAs were identified that are either specifically (PRP) or preferentially (GBSS and CHS) present in the boll wall cells. Therefore, the promoters driving expression of these transcripts were targeted for isolation. This chapter describes the isolation and characterisation of these three promoters.

Gene expression in plants is spatially and temporally regulated by the binding of specific transcription factors to *cis*-acting regulatory elements, known as CAREs, normally found within the promoter regions of genes. These CAREs are generally sequence motifs of between five and 20 nucleotides in length and are usually located near the 5' end of genes (Rombauts *et al.*, 2003). The CAREs are typically located upstream of the transcription start site, but regulatory elements can also be located downstream, for example in the first intron (Reddy and Reddy, 2004).

A promoter can be divided into two sections known as the proximal and distal regions (Rombauts *et al.*, 2003). The proximal, or core, region near the start of transcription is believed to direct basal levels of transcription via assembly of the RNA polymerase II complex at the correct position (Berk, 1999). Promoter elements within the proximal promoter include the TATA box, which binds TATA-binding protein, and elements such as the CAAT box that bind general transcription factors (Featherstone, 2002). The distal part of the promoter directs transcription above basal levels, with CAREs in this region binding various transcription factors that regulate spatial and temporal gene expression (Fessele *et al.*, 2002).

The 5' ends of plant promoters are often poorly defined and can be variable in length (Rombauts *et al.*, 2003). Generally, major promoter elements are located with 2 kb of the transcriptional start site of genes (Chandler and Vaucheret, 2001, Tian *et al.*, 2004). Promoter regions in the range of several hundred base pairs are usually sufficient to produce expression patterns of reporter genes that mirror those seen for the endogenous gene *in planta* (Hsu *et al.*,

1999, Sunilkumar *et al.*, 2002b, Yamagata *et al.*, 2002, Harmer, 2003, Agius *et al.*, 2005, Delaney, 2005). In addition to the proximal and distal promoter regions, more distant regulatory regions can act as enhancer or repressor elements and may be located several kilobases away from the proximal promoter region (Engstrom *et al.*, 2004, Tian *et al.*, 2004). Transcription can also be influenced by epigenetic mechanisms that influence the accessibility of promoter regions to transcription factors. These mechanisms include chromatin packaging and associated histone methylation and/or acetylation as well as methylation of promoter regions (Meyer, 2000, Bender, 2004, Steimer *et al.*, 2004, Tian *et al.*, 2005).

4.2: Genomic analysis

To determine how many sequences similar to the PRP, GBSS and CHS cDNAs exist in the genome of *G. hirsutum*, and thus whether a PCR-based approach to isolating the promoters would be feasible, the gene family complexities were examined using genomic Southern analysis. Genomic DNA was digested with several restriction enzymes that have six base pair recognition sites (*Bam*HI, *Eco*RI, *Eco*RV, *Hin*dIII, *Nco*I, *Sac*I and *Xba*I). Southern blots were hybridised with labelled cDNA probes to determine the number of hybridising fragments to estimate genomic copy number (figure 4.1). Prior to Southern analysis, the location of restriction enzyme recognition sites within the cDNA sequences was determined, to aid interpretation of the Southern autoradiographs.

4.2.1: Proline rich protein Southern analysis

The genomic Southern hybridised with the labelled PRP cDNA showed a low number of hybridising bands. The *Bam*HI, *Eco*RI and *Xba*I lanes demonstrated hybridisation to single fragments of 3.4 kb, 6.8 kb and 6 kb, respectively, while the *Hin*dIII lane showed hybridisation to three smaller fragments of 0.7 kb, 2.4 kb and 2.7 kb and the *Eco*RV, *Nco*I and *Sac*I lanes revealed no prominent bands (figure 4.1). Within the PRP cDNA cloned sequence, single restriction enzyme recognition sites corresponding to *Bam*HI, *Hin*dIII, *Sac*I and *Xba*I were identified. The presence of several hybridising bands in the *Hin*dIII lane can therefore be partially explained by the presence of an enzyme recognition site within the cDNA allowing for the possible hybridisation of the probe to two fragments. The overall hybridisation pattern of the PRP probe suggests that there are one or two genomic copies of the gene. It is also possible that allelic variation of a single gene could give the observed hybridisation pattern. Some minor bands were also observed at 1.5 kb (*Hin*dIII), 1.3 kb (*Nco*I) and 8.5 kb (*Xba*I), which suggest the presence of genomic sequences encoding other similar proline rich proteins.

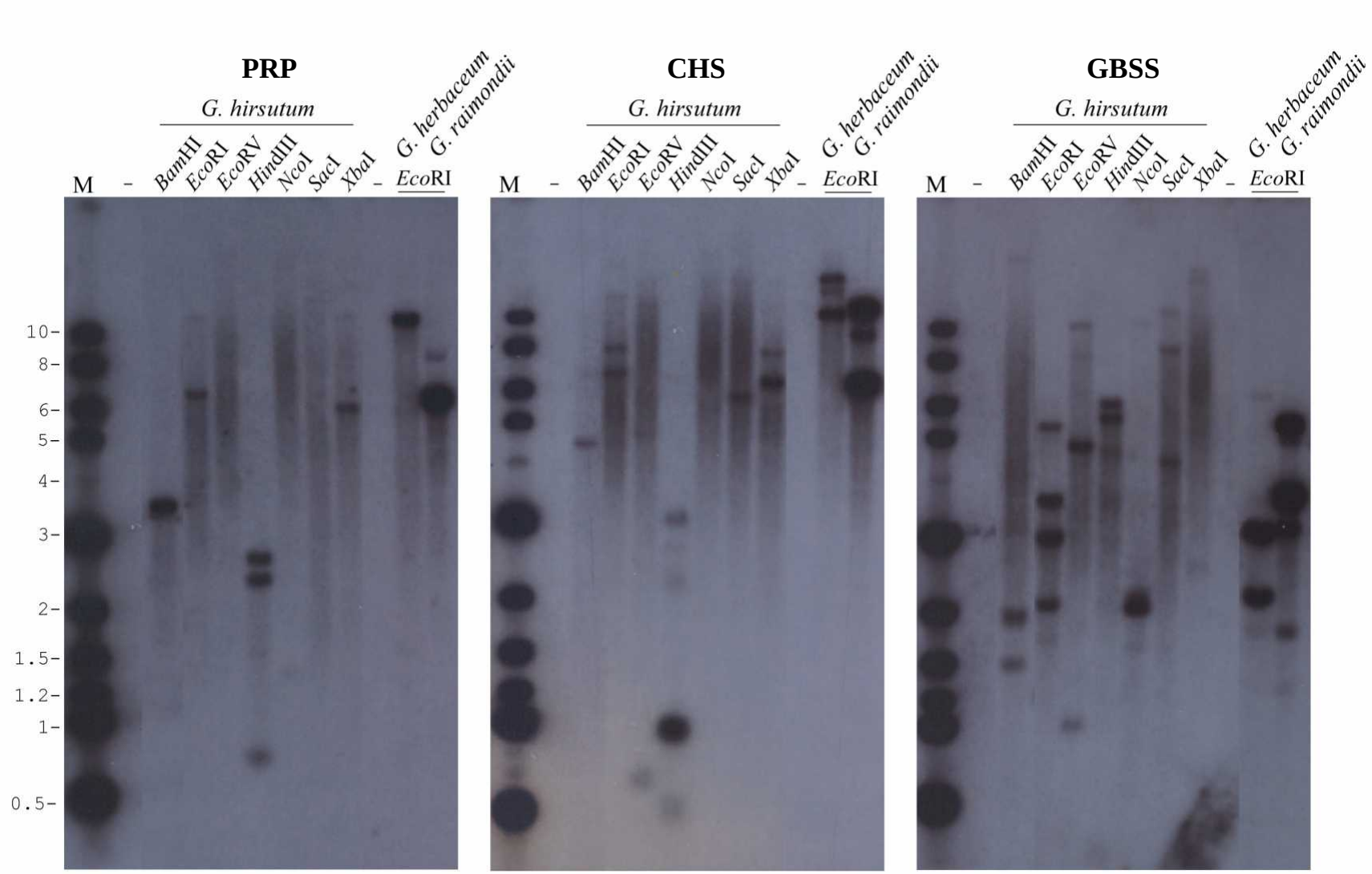


Figure 4.1: Genomic Southern analysis. Genomic Southerns were hybridised with the PRP, CHS and GBSS cDNAs as indicated. Each lane contains 10 μg of genomic DNA (*G. hirsutum, G. herbaceum* or *G. raimondii*), digested with restriction enzymes as shown. Genomic DNA from the extant diploid ancestors are shown in the two far right lanes, digested with *Eco*RI. Molecular weight markers sizes are indicated in kb.

4.2.2: Chalcone synthase Southern analysis

The hybridisation pattern of the genomic Southern hybridised with the CHS cDNA probe was more complex. Signals can be seen in most lanes at various sizes including 4.2 kb (*Bam*HI), 6.5 kb and 7.8 kb (*Eco*RI), 0.7 kb (*Eco*RV), 2.9 kb and 0.9 kb (*Hin*dIII), 5.7 kb (*Sac*I) and 6.0 kb and 7.2 kb (*Xba*I) (figure 4.1). There are also numerous other more faintly hybridising fragments in most lanes. The CHS cDNA contained single recognition sites for *Eco*RV and *Sac*I and two recognition sites for *Hin*dIII. These two *Hin*dIII sites lie 884 bp apart, consistent with the ~0.9 kb band observed in the *Hin*dIII lane. The presence of two or three hybridising bands in most lanes suggests that a small CHS gene family exists in cotton. This is in agreement with other published reports, which suggest that most plant species contain between one and three copies of the gene (Junghans *et al.*, 1993). The large number of weakly hybridising bands suggests that there may be other similar genomic sequences within the *G. hirsutum* genome. It is also possible that some of the larger bands may represent hybridisation to genomic DNA that has not been completely digested.

4.2.3: Granule-bound starch synthase Southern analysis

The hybridisation pattern of the genomic Southern that was hybridised with the GBSS cDNA probe was the most complex. Multiple prominent bands were observed in most lanes including *Bam*HI (1.5 kb and 2 kb), *Eco*RI (2.1 kb, 3.0 kb, 3.5 kb and 5.3 kb), *Eco*RV (3.8 kb), *Hin*dIII (4.5 kb and 4.9 kb), *Nco*I (2 kb) and *Sac*I (4.4 kb) (figure 4.1). The GBSS probe also hybridised to several fragments of varying sizes in most lanes. Analysis of the cDNA sequence revealed recognition sites for *Bam*HI, *Eco*RV, *Hin*dIII, and *Sac*I located within the GBSS sequence, and two *Nco*I sites located 1,268 bp apart. There is no hybridising fragment in the *Nco*I lane at a corresponding size, with the only major band of hybridisation at 2.0 kb in size. This suggests that the genomic sequence may contain an intron in the region between the two *Nco*I sites. The presence of several strongly hybridising bands in most lanes indicates that there are probably at least two or three highly similar genomic sequences. The numerous other, more weakly hybridising bands indicate that there are several highly similar genomic regions.

4.2.4: Genomic origins

The genomic origins of the genes that correspond to each of the three cDNAs were assessed by genomic Southern analysis. The cDNA probes were hybridised to *Eco*RI digested-genomic DNA from the candidate ancestral diploid cotton species *G. herbaceum* and *G. raimondii* (figure 4.1). All three of the labelled cDNA probes hybridised to the digested DNA of both diploid species with one prominent fragment observed for the PRP probe and

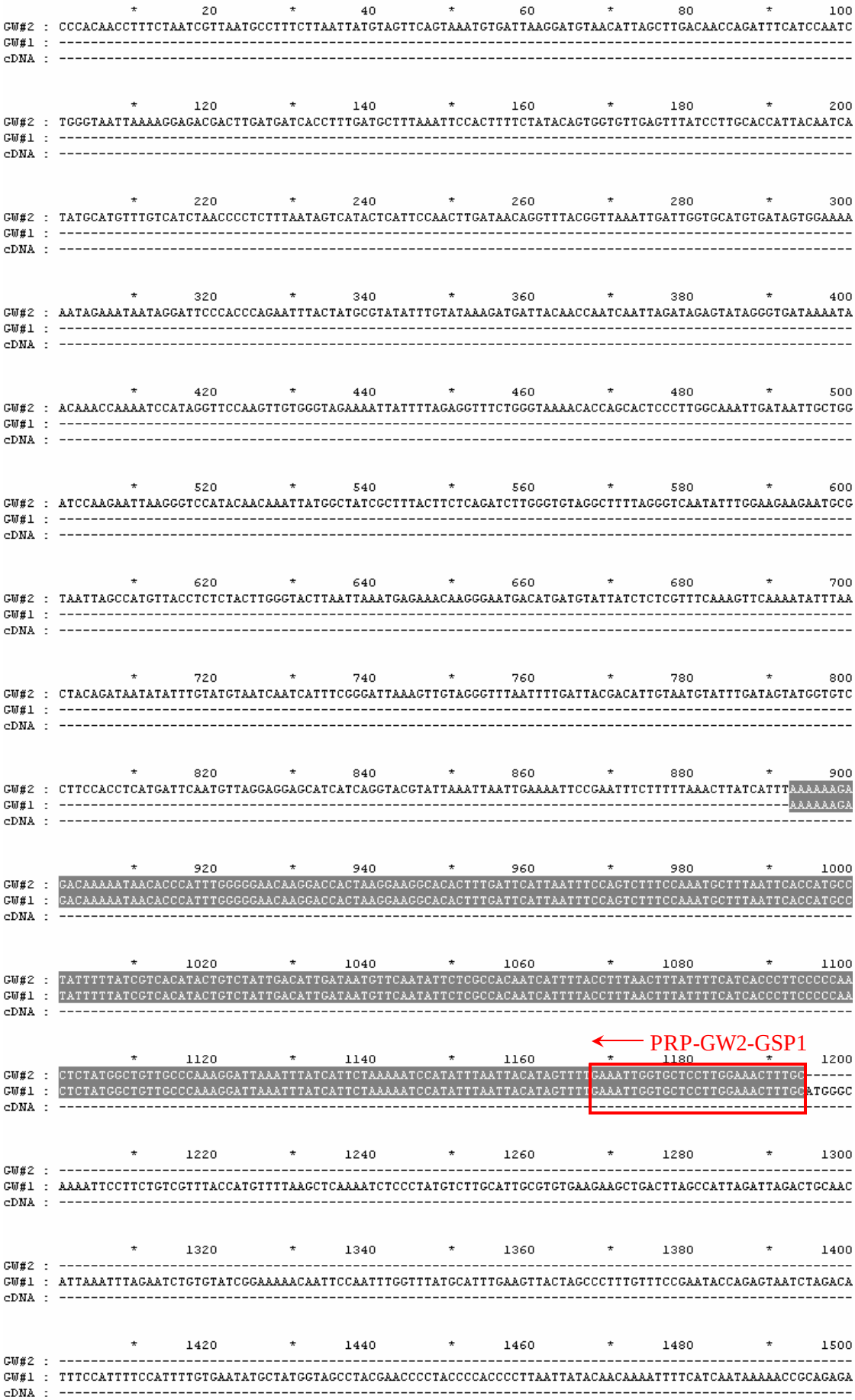
two prominent fragments observed for the CHS and GBSS probes. This indicates that both of the ancestral cotton diploid species probably contained corresponding genes. This is to be expected for CHS and GBSS as they play important roles in most plant species. In the case of PRP, the suggested role for this gene product in boll wall growth and development may explain its presence in the ancestral diploids.

4.3: Promoter identification

Analysis of the family size of the genes corresponding to the three identified cDNAs indicates that each gene is probably represented by between one and four genomic regions validating a PCR-based approach to isolating the promoters. The approach selected was a PCR-based genome walking method, Genome Walker® PCR. This method utilises PCR with gene specific primers to amplify upstream promoter sequence that overlap with the known cDNA sequence. DNA 'libraries' are used as PCR templates and are prepared by digestion of genomic DNA with different restriction enzymes (*Dra*I, *Eco*RV, *Pvu*II, *Sca*I, *Stu*I, *Sma*I, *Ssp*I, *Hae*III, *Alu*I, *Rga*I), followed by ligation of adaptors to both ends of the genomic fragments. Gene-specific and adaptor-specific primers are used for primary, followed by secondary (or nested) PCR to ensure that the desired promoter sequence is amplified. Usually, a set of ten PCRs is performed on the ten restricted DNA libraries, followed by a set of ten nested PCRs on the products generated by the primary PCR. The products are then cloned, sequenced and compared with the overlapping cDNA sequence. It is important to compare the generated promoter fragment with the 5' cDNA end to ensure that the correct genomic region has been isolated, as the PCR-based Genome Walker® method can amplify similar non-target sequences. In cases where more promoter sequence is required than is obtained by this method, a second application of Genome Walker® PCR can be performed with new gene-specific primers that bind further upstream than the primers used in the first Genome Walker® PCRs.

4.3.1: Proline rich protein promoter identification

Primers were designed to the cDNA sequence of the PRP such that there would be 145 bp of overlap between the cDNA sequence and the amplified promoter sequence (figure 4.2). After primary and subsequent nested PCR, four PCR products of greater than 700 bp were cloned and sequenced (figure 4.3B). These four PCR products were derived from lane one (product #1), lane five (product #5) and two bands from lane seven (#7-upper and #7-lower). Upon alignment with the previously characterised PRP cDNA, 145 bp of overlapping sequence was apparent (figure 4.4). The four PCR products may be assigned to two groups based on homology with each other and with the PRP cDNA sequence (figure 4.4). Products


PRP-GW2-GSP1

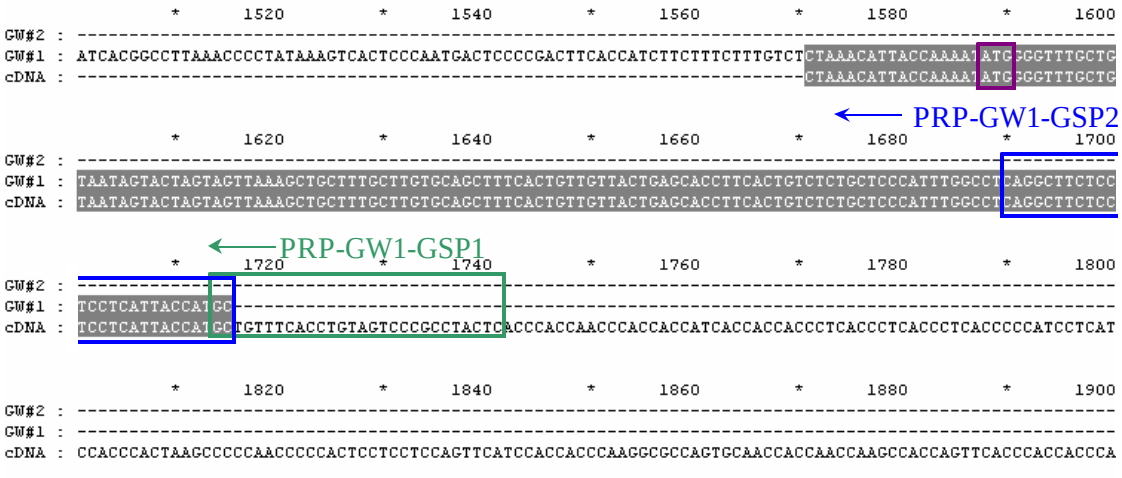


Figure 4.2: Proline rich protein Genome Walker® PCR primers and sequences.
Sequences of the 5' region of the cDNA identified in the differential screen (cDNA), the promoter sequence identified in the first Genome Walker® PCR experiment (GW#1) and the promoter sequence identified in the second Genome Walker® PCR experiment (GW#2) are shown in the alignment. The putative ATG start site of translation is indicated with a purple box and the binding sites of the primers used for Genome Walker® PCRs are indicated with boxes (labelled as Proline-Rich Protein-Genome Walk 1 or 2-Gene Specific Primer 1 or 2).

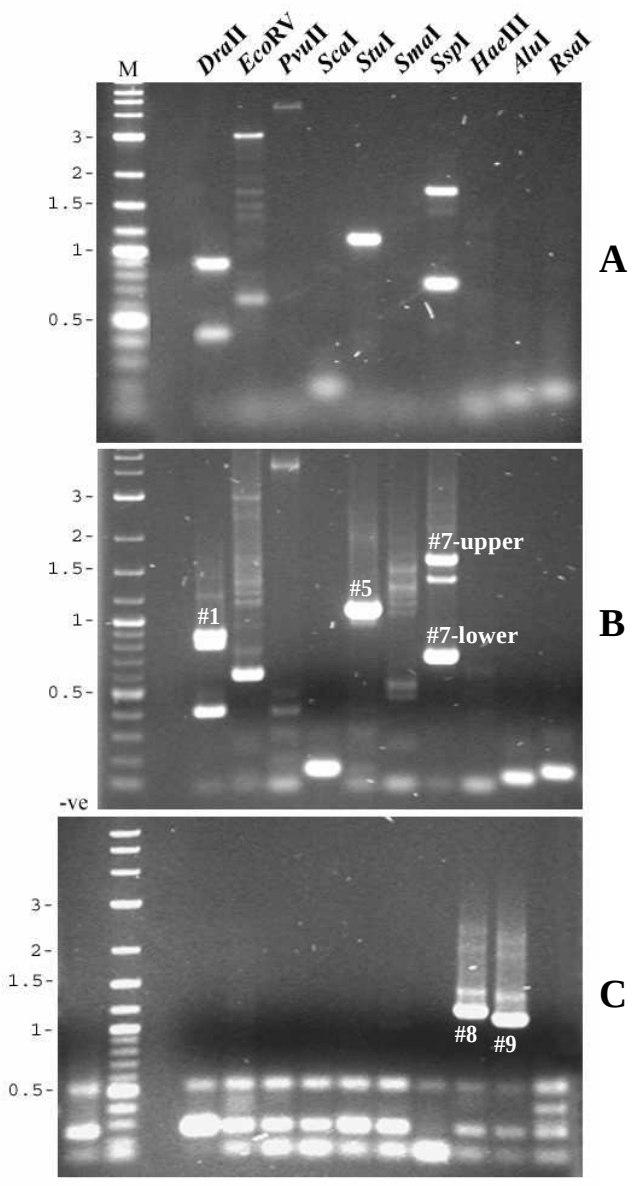


Figure 4.3: Proline rich protein Genome Walker® PCR products.
Products from the PRP Genome Walker® PCR reactions are shown, with the enzyme used to generate each of the ten Genome Walker® libraries indicated above the corresponding lane. DNA ladder sizes are shown in kb pairs and Genome Walker® PCR products discussed in the text are labelled. Two rounds of Genome Walker® PCR were performed to identify the PRP promoter, with the products of the primary and secondary PCRs of the first 'walk' shown in panels A and B, respectively. The products of the primary PCR of the second 'walk' are shown in panel C. The Genome Walker® PCR products from the nested PCR of the second 'walk' contained several bands between 100 and 500 bp in size, which were also observed in the negative control lane ("-ve" - no DNA template), indicating the presence of contamination or the formation of PCR products from primer dimerisation.

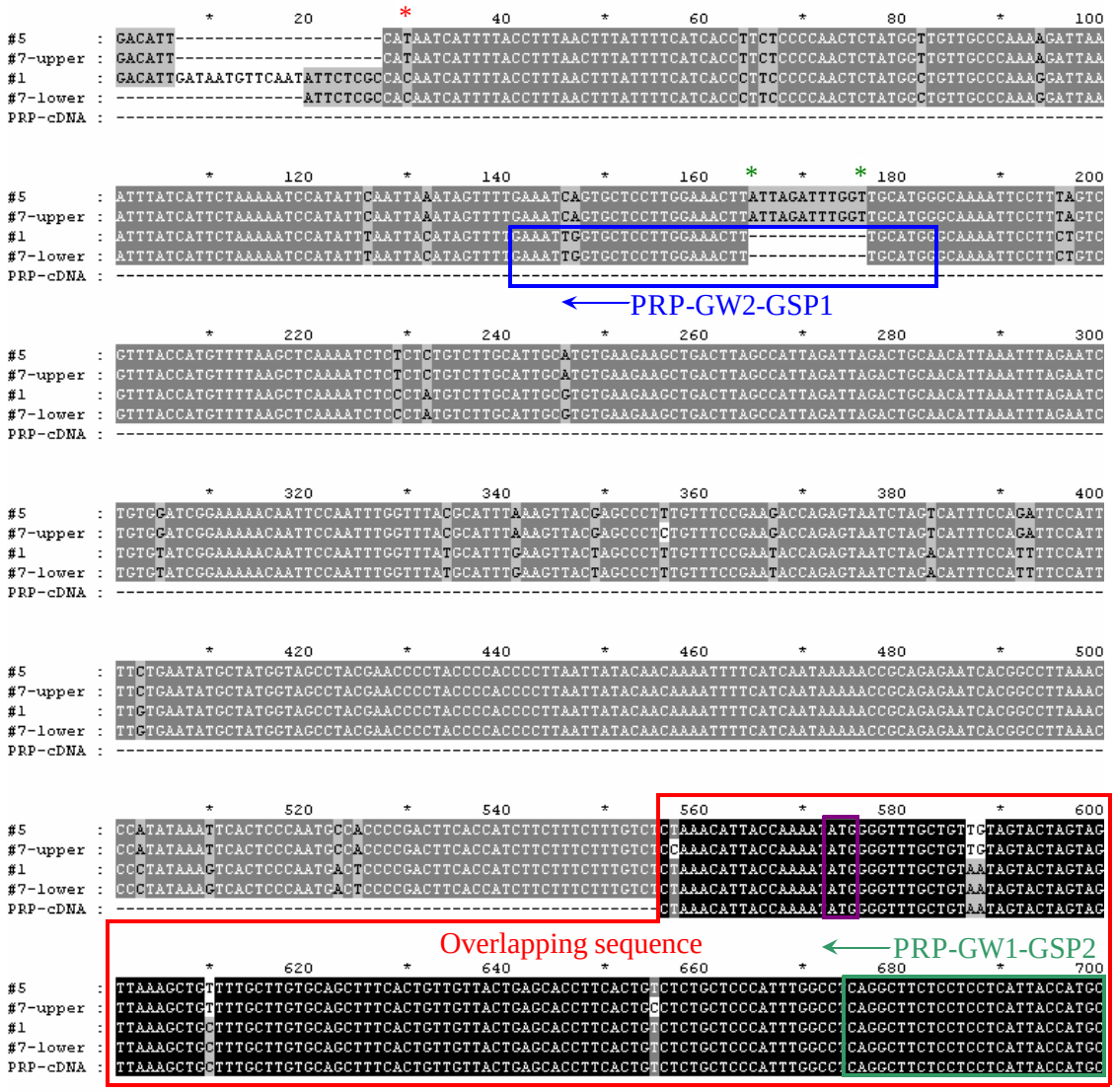


Figure 4.4: Proline rich protein Genome Walker® PCR product alignment.
An alignment of the products obtained from the first PRP Genome Walker® PCR experiment and the 5' region of the PRP cDNA is shown. A large red box indicates the region of sequence overlap between the PRP cDNA and the Genome Walker® PCR products. The ATG start site of translation is indicated with a purple box and the binding sites of the primers used for Genome Walker® PCRs are indicated with boxes. Red and green asterisks indicate the nucleotide positions discussed in chapter four.

#1 and #7-lower are identical to the 5' end of the PRP cDNA, while products #5 and #7-upper contain three and five nucleotide differences respectively, with three of the differences shared between the two sequences. Products #1 and #7-lower are identical to each other, while #5 and #7-upper are 99.6% similar. There are 31 nucleotide positions in the alignment where the two groups differ, for example at nucleotide 30 where #1 and #7-lower contain a cytosine and #5 and #7-upper contain a thymine (figure 4.4, red asterisk). As well as these single nucleotide differences, there are three regions of insertion/deletion between the two groups, such as between nucleotides positions 165 and 176 (figure 4.4, green asterisk). The similarity of products #1 and #7-lower with the PRP promoter suggest that they correspond to this genomic region while products #5 and #7-upper may correspond to a different genomic region. Product #1 is the longer of the two clones that correspond to the PRP cDNA, and contains 678 bp of novel upstream genomic sequence.

Because this region is relatively short, primers complementary to this putative promoter region were designed for use in a second Genome Walker® PCR experiment to extend the region of identified promoter. One of the primers (PRP-GW2-GSP1) was designed to span a region of difference between the two classes of sequence described previously in order to selectively amplify products corresponding the correct genomic region (figures 4.2 and 4.4). Primary PCR using the ten template libraries resulted in prominent bands in lanes eight and nine, termed products #8 and #9 (figure 4.3C). Product #8 was longer and its sequence was identical to the previously identified fragment in the overlapping region. The homology of this product with the previously identified promoter fragment suggests that it corresponds to the boll wall PRP promoter. When the two promoter fragments generated by Genome Walker® PCR were computationally compiled, a total of 1587 bp of sequence upstream of the start site of translation were obtained.

4.3.2: Granule-bound starch synthase promoter identification

Primary and nested PCR were performed as above with products from lane one (#1-upper and #1-lower), two (#2), four (#4) and seven (#7) cloned and sequenced (figures 4.5 and 4.6B). When the sequence of these four Genome Walker® PCR products were aligned with the GBSS cDNA, a region of strong homology (160 bp) was observed followed by a region with no substantial homology (103 bp) (figure 4.7). The sequence differences between the cDNA sequence and the Genome Walker® PCR products suggest that the products do not correspond to the desired promoter. This is not surprising as, particularly in cotton, Genome Walker® PCR often amplifies genomic sequences that are very similar to the desired region due to the PCR-based nature of the technique. Therefore, a primer was designed to select

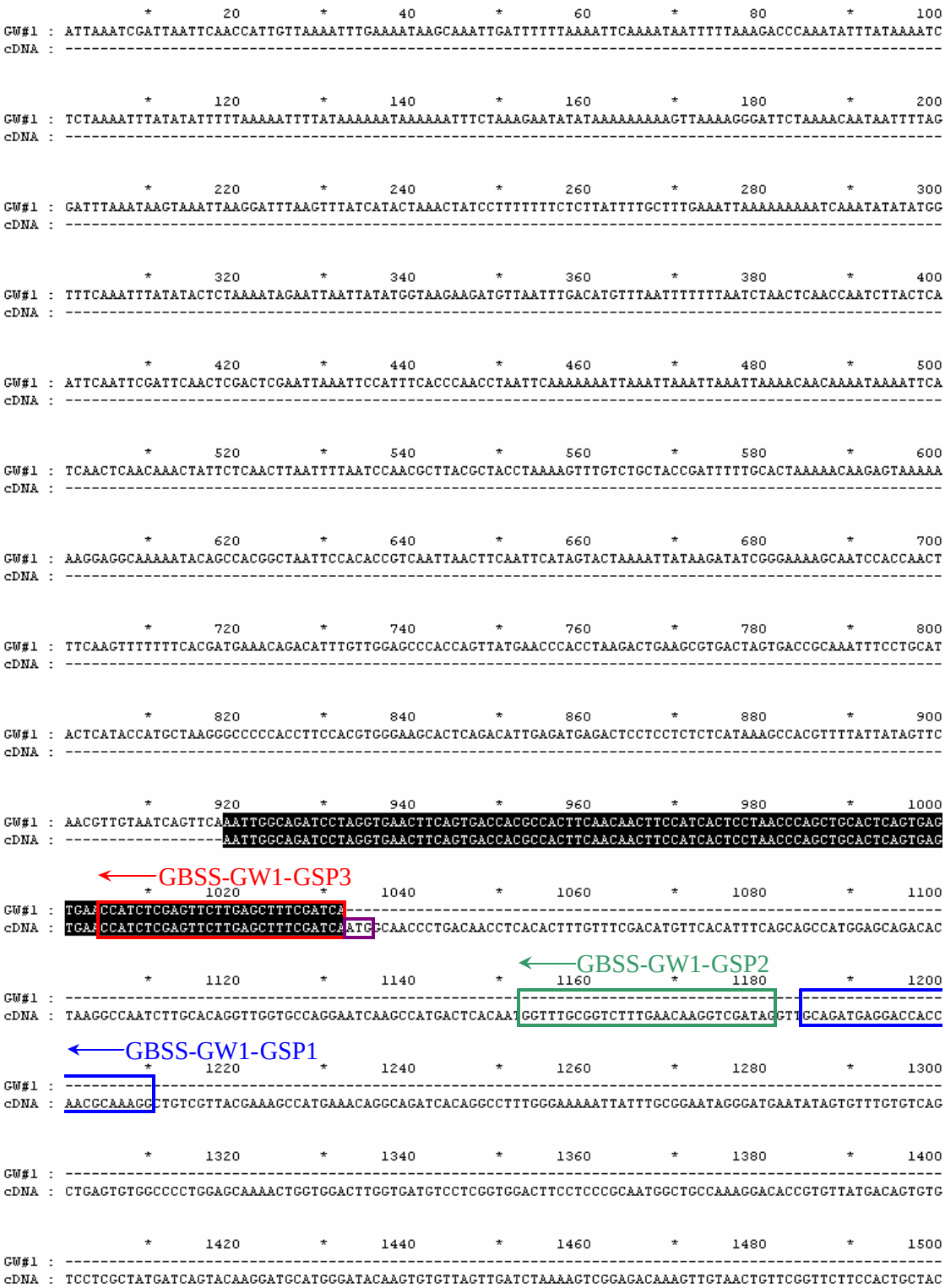


Figure 4.5: Granule-bound starch synthase Genome Walker® PCR primers and sequences. Sequences of the 5' region of the cDNA identified in the differential screen (cDNA) and the promoter sequence identified in the first Genome Walker® PCR experiment (GW#1) are shown in the alignment. The putative ATG start site of translation is indicated with a purple box and the binding sites of the primers used for Genome Walker® PCRs are indicated with boxes (labelled as Granule-Bound Starch Synthase-Genome Walk 1-Gene Specific Primer 1, 2 or 3).

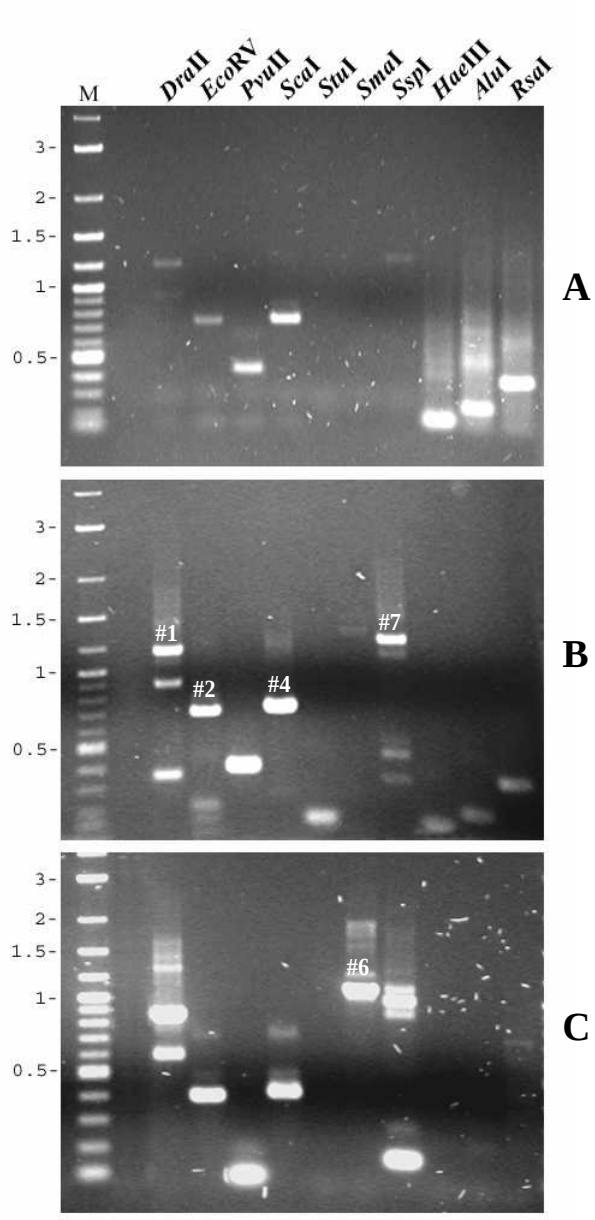


Figure 4.6: Granule-bound starch synthase Genome Walker® PCR products.
Products from the GBSS Genome Walker® PCR reactions are shown, with the enzyme used to generate each of the ten Genome Walker® libraries indicated above the corresponding lane. DNA ladder sizes are shown in kb pairs and Genome Walker® PCR products discussed in the text are labelled. One round of Genome Walker® PCR were performed to identify the GBSS promoter, with the primary, secondary and tertiary PCR reaction products shown in panels A, B and C, respectively.

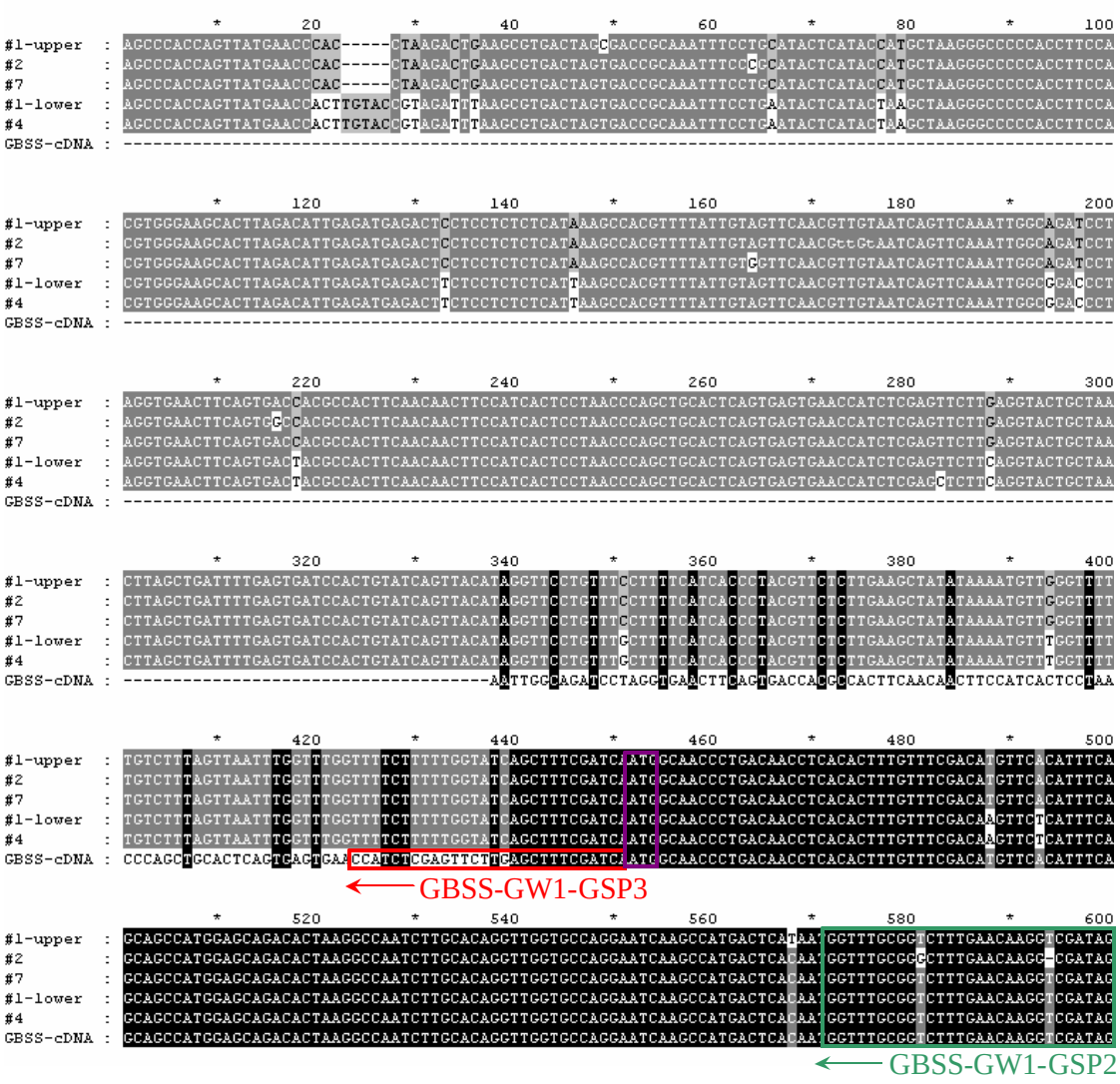


Figure 4.7: Granule-bound starch synthase Genome Walker® PCR product alignment. An alignment of the products obtained from the GBSS Genome Walker® PCR experiment (#1-upper, #2, #7, #1-lower and #4) and the 5' region of the GBSS cDNA (GBSS-cDNA) is shown. The binding sites of the primers used for this Genome Walker® PCR experiment are indicated with a red box, and the ATG start site of translation is shown with a purple box.

against the amplification of these none-target sequences (figure 4.7) and a second round of nested PCR was performed (figure 4.6C). The products of this amplification correlated in size with the products of the first nested PCR except for that in lane six (#6). This product was cloned and sequenced and showed 100% homology within the region of overlap with the cDNA (figure 4.5). The clone contains 918 bp of novel GBSS upstream sequence that, in addition to the 114 bp of 5' UTR present within the isolated cDNA, totals 1,032 bp prior to the start site of translation initiation. No further promoter sequence was isolated, as the 5' 600 bp was 80% A/T, with extensive stretches of adenine and thymine, making primer design in this region difficult.

4.3.3: Chalcone synthase promoter identification

To amplify the promoter region corresponding to the CHS cDNA, primers were designed to create a 170 bp sequence overlap between the 5' cDNA end and potential Genome Walker® PCR products (figure 4.8). Amplification yielded several clear PCR bands (figure 4.9A). The longest product was observed in lane seven (#7), and this fragment is identical to the overlapping 5' cDNA sequence (figure 4.8). This clone contains only a small fragment (590 bp), containing 496 bp upstream of the start site of translation (figure 4.8). A second round of Genome Walker® PCR was performed with primary and nested PCR (figure 4.8) and fragments greater than 400 bp in size were cloned and sequenced from lanes one (#1), three (#3), five (#5) and eight (#8) (figure 4.9C). Product #8 showed no homology to the known CHS cDNA sequence outside of the primer region, while the other PCR products showed strong homology (encompassing ~250 bp) to the CHS promoter sequence (figure 4.10). Products #3 and #5 were the longest, containing 535 bp and 520 bp of novel promoter, respectively. There are four nucleotide differences between the two sequences with three nucleotide substitutions and one deletion/insertion (figure 4.10). Together the products provide a total of 723 bp of sequence upstream of the start site of translation. No further CHS promoter sequence was obtained because of the A/T rich nature of the promoter and the repetitive nature of the sequence in this region.

4.3.4: Genome Walker® PCR error rates

Some of the products obtained from the Genome Walker® PCRs for each of the promoters contained sequence differences when compared to each other and to the 5' cDNA sequence, which are likely to have resulted from *in vitro* polymerase errors or cloning artifacts. For the Genome Walker® PCRs, a mixture of two polymerases was used; the non-proofreading DNA polymerase enzyme from *Thermus aquaticus* (*Taq*) and the proofreading DNA polymerase from *Pyrococcus furiosus* (*Pfu*) at a molar ratio of 16:1. For the polymerase

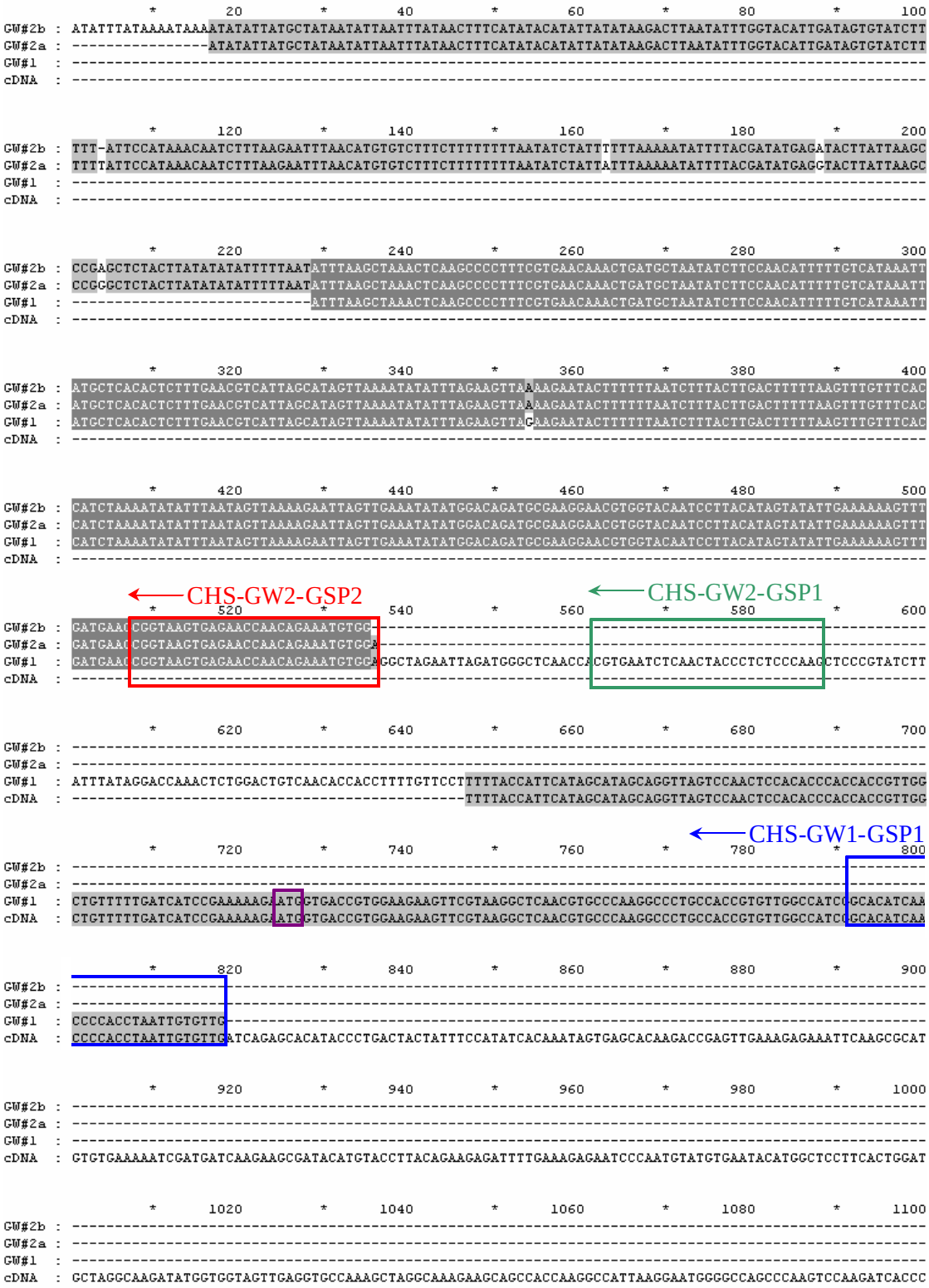


Figure 4.8: Chalcone synthase Genome Walker® PCR primers and sequences.
Sequences of the 5' region of the cDNA identified in the differential screen (cDNA), the promoter sequence identified in the first Genome Walker® PCR experiment (GW#1) and the two different Genome Walker® PCR products identified in the second Genome Walker® PCR experiment (GW#2a and GW#2b) are shown in the alignment. The putative ATG start site of translation is indicated with a purple box and the binding sites of the primers used for Genome Walker® PCRs are indicated with boxes (labelled as Chalcone Synthase-Genome Walk 1 or 2-Gene Specific Primer 1 or 2).

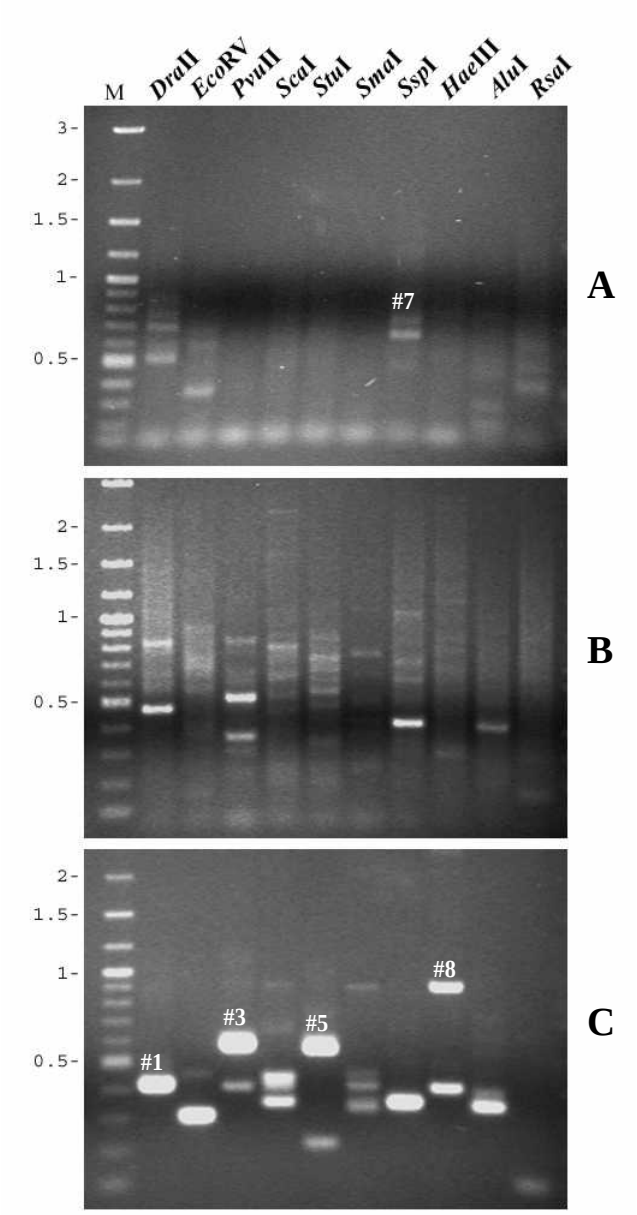


Figure 4.9: Chalcone synthase Genome Walker® PCR products.
Products from the CHS Genome Walker® PCR reactions are shown, with the enzyme used to generate each of the ten Genome Walker® libraries indicated above the corresponding lane. DNA ladder sizes are shown in kb pairs and Genome Walker® PCR products discussed in the text are labelled. Two rounds of Genome Walker® PCR were performed to identify the CHS promoter, with the primary PCR reaction products from the first 'walk' shown in panel A and the products of the primary and nested PCR reactions of the second 'walk' shown in panel B and C, respectively.

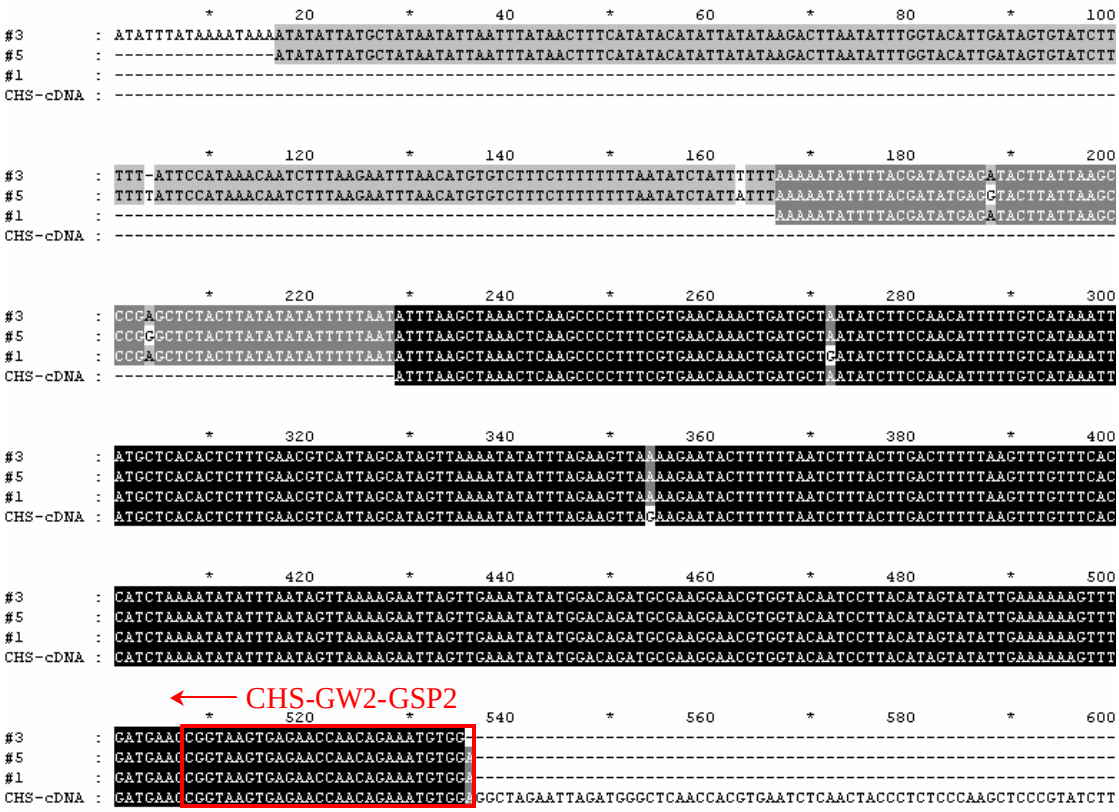


Figure 4.10: Chalcone synthase Genome Walker® PCR product alignment.
An alignment of the products obtained from the second CHS Genome Walker® PCR experiment (#1, #5 and #3) and the 5' region of the CHS cDNA (CHS-cDNA) is shown. The binding site of the primer used for this Genome Walker® PCR is indicated with a red box.

mixture and conditions used (chapter 2), there is a 7.6 x 10^{-7} chance of a mismatched base being incorporated into the amplified product per base pair per duplication (Cline *et al.*, 1996). For a 1 kb stretch of promoter, amplified through primary and secondary Genome Walker® PCR (a total of 50 cycles), there is a 38% chance of the product containing one mutation. The occurrence of sequence differences within the Genome Walker® PCR products is consistent with this predicted rate of mutation.

4.4: Promoter isolation and cloning into transient transformation reporter vector

The promoter elements that are responsible for the transcription of a gene are usually located close to the 5' end of the gene, and hence the regions identified here are likely to contain the elements that drive high levels of transcription of the corresponding mRNAs in the boll wall (Tyagi, 2001). Transient transformations of cotton tissues with each of the promoters driving reporter gene expression were performed to test this hypothesis. The promoters were cloned into the promoterless pJK.KiwiGUS.ocs.Kmf(-) (pJK) vector (Kirschman and Cramer, 1988), upstream of the reporter gene. The reporter gene used in this study is the *gusA* gene, which encodes the β-glucuronidase (GUS) enzyme. The β-glucuronidase protein cleaves a substrate (X-GLUC), which is introduced to the tissue during the staining procedure, to produce a blue precipitate.

In order to assemble each of the promoters into a continuous stretch of DNA for cloning, primers were designed close to the 5' and 3' ends of the promoter sequences (figures 4.11, 4.12 and 4.13) and PCR using genomic DNA as template was performed. In order to generate functional expression clones, the ATG start site of translation was not included in the amplified promoters because the *gusA* gene in the vector contains an ATG translational start site. The primers were designed with restriction enzyme recognition sites at their 5' ends to allow cloning into the multiple cloning site of the reporter vector.

Promoters were amplified from genomic DNA with either a proofreading (*Pfu*) or a non-proofreading polymerase (*Taq*). Amplification using proofreading enzymes is the more straightforward method but generally polymerases with increased proofreading capability possess lower processivities, making amplification of target sequences difficult. Therefore, promoters that were difficult to amplify with *Pfu* polymerase were amplified with *Taq* polymerase. When *Taq* polymerase was used for promoter amplification, the PCR products from several independent reactions were sequenced and analysed to determine which products

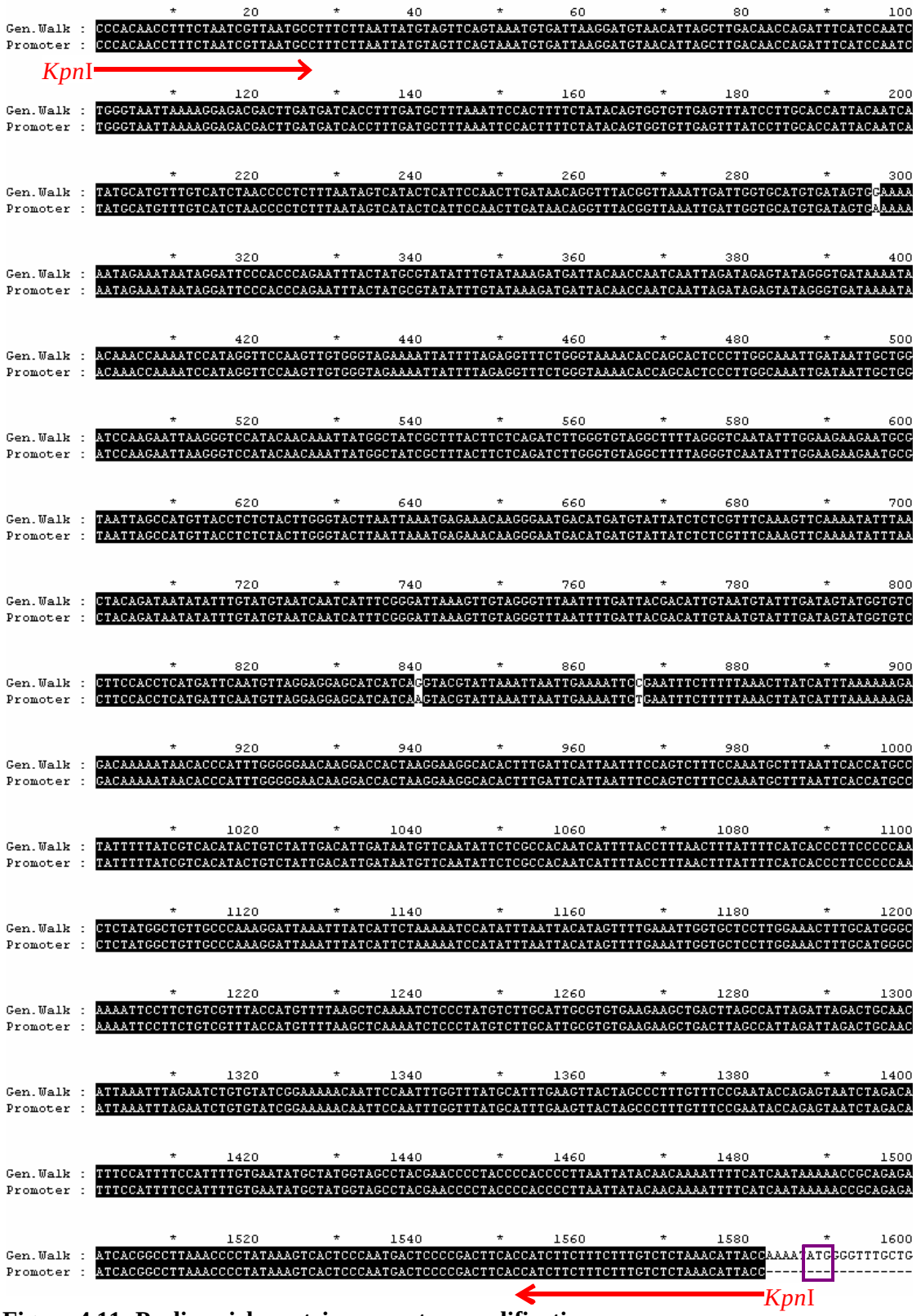


Figure 4.11: Proline rich protein promoter amplification.
An alignment of the computationally compiled Genome Walker® PCR products (Gen.Walk.) and the promoter generated by genomic PCR (Promoter) is shown. The putative ATG start site of translation is indicated with a purple box and the binding sites of the primers used for PCR amplification of the promoter are indicated with red arrows, with the included restriction enzyme recognition sites indicated in red.

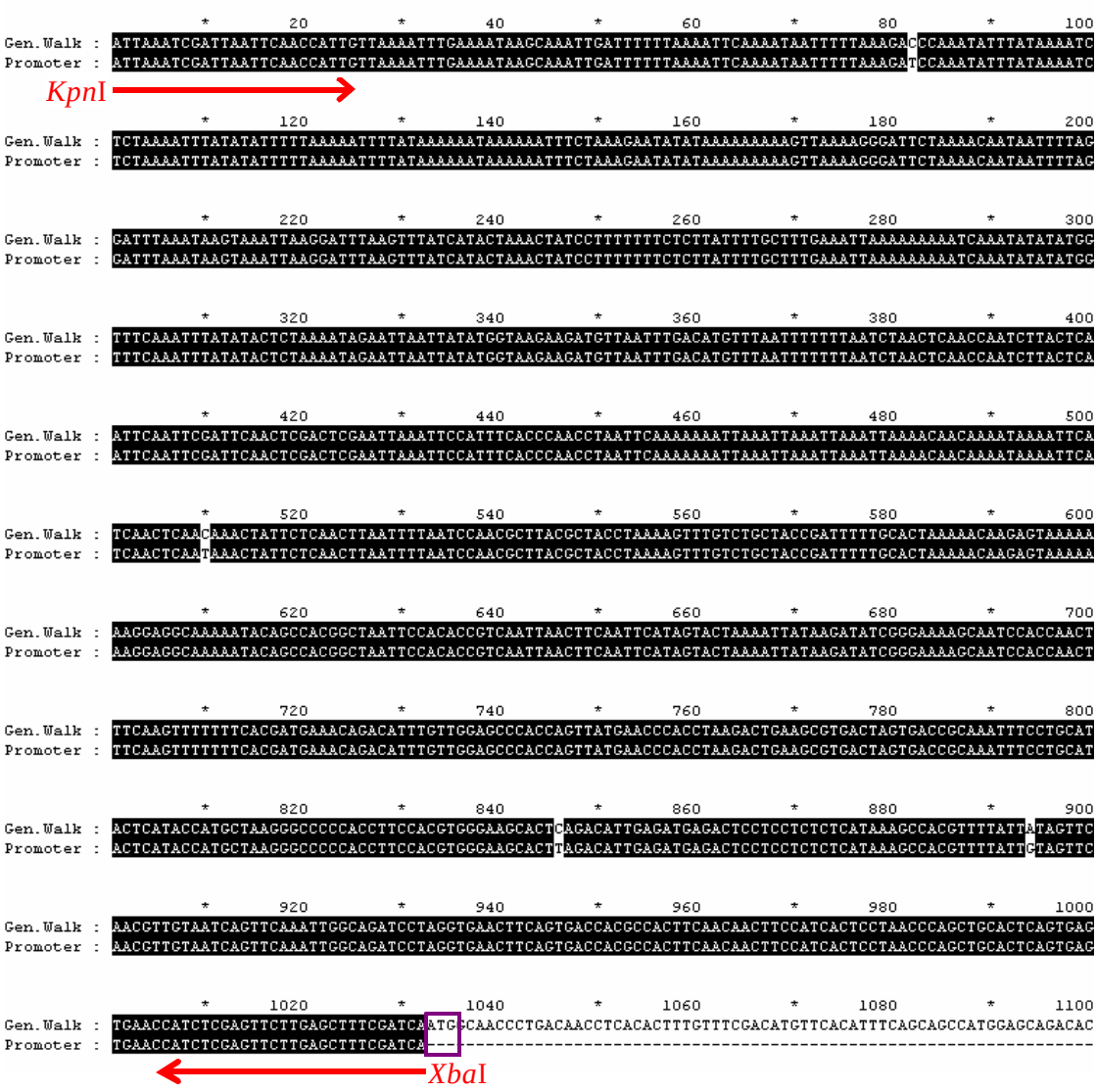


Figure 4.12: Granule-bound starch synthase promoter amplification.
An alignment of the Genome Walker® PCR promoter (Gen.Walk.) and the promoter generated by genomic PCR (Promoter) is shown. The putative ATG start site of translation is indicated with a purple box and the binding sites of the primers used for PCR amplification of the promoter are indicated with red arrows with the included restriction enzyme recognition sites indicated in red.

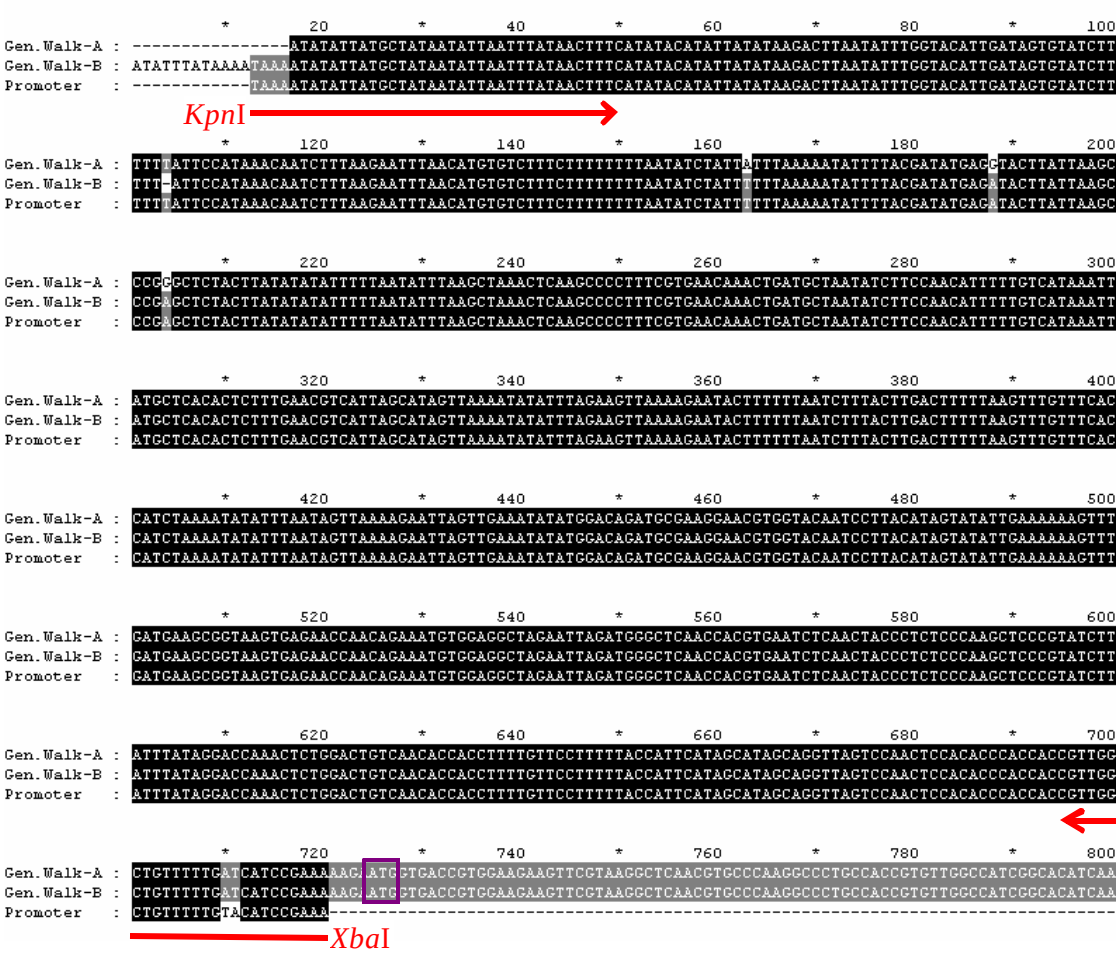


Figure 4.13: Chalcone synthase promoter amplification.
An alignment of the computationally compiled Genome Walker® PCR products (Gen.Walk.) and the promoter generated by genomic PCR (Promoter) is shown. The putative ATG start site of translation is indicated with a purple box and the binding sites of the primers used for PCR amplification of the promoter are indicated with red arrows, with the included restriction enzyme recognition sites indicated in red.

represent the correct promoter sequence. This approach assumes that the same mutation does not reproducibly occur at a single nucleotide position in independent PCRs.

The CHS and GBSS promoters were both amplified from genomic DNA with *Taq* polymerase using primers that contain *Kpn*I and *Xba*I restriction enzyme recognition sites (figures 4.12 and 4.13). For each promoter, PCR products were cloned from up to seven independent PCR reactions and sequenced, with consensus promoter sequences determined (figure 4.14).

The PRP promoter was initially amplified from genomic DNA using *Pfu* polymerase and cloned into the Gateway vector pMDC162GUS (Curtis and Grossniklaus, 2003). Several independent PCR products were sequenced and a clone that matched the consensus sequence of the independent PCRs was selected. From this clone, PCR was performed, using primers with *Kpn*I restriction enzyme recognition sites (figure 4.11), and the products were digested and cloned into pJK.

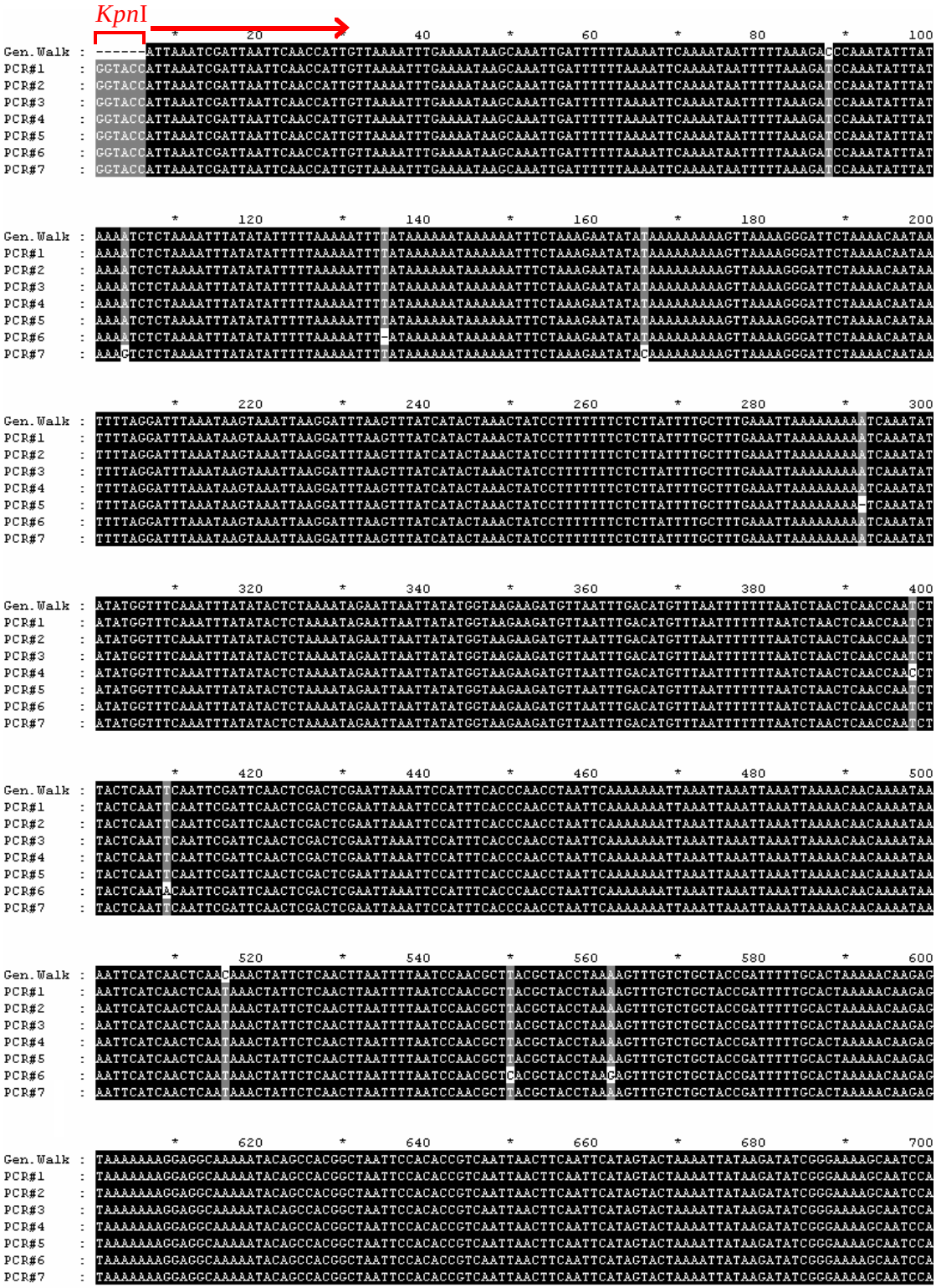
A positive control promoter, *CaMV 35S*, which is known to drive strong constitutive expression (Odell *et al.*, 1985, Jefferson *et al.*, 1987) was utilised in this study. The promoter was obtained by PCR, using primers with *Kpn*I restriction enzyme recognition sites, from pMDC43GFP (Curtis and Grossniklaus, 2003) using *Taq* polymerase. PCR products were cloned into pGEM-T Easy and then transferred into pJK by restriction digest and ligation.

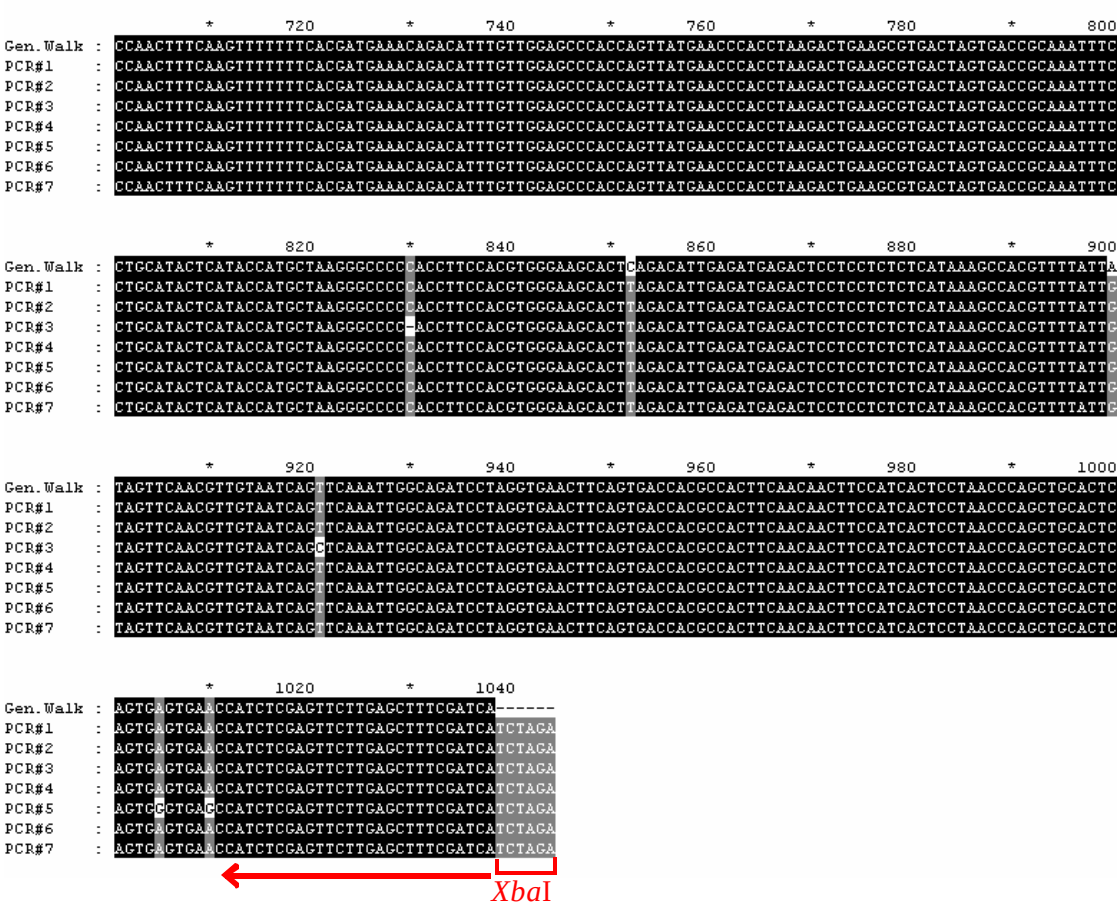
When the promoter sequences that were generated by genomic PCR were compared to the promoter sequence generated by Genome Walker® PCR (figures 4.11, 4.12 and 4.13), several nucleotide differences were seen within the three promoters. As previously discussed, the nucleotide differences are probably due to errors introduced during Genome Walker® PCR, and thus the true promoter sequences are likely to be those generated by direct genomic PCR.

4.5: Summary and conclusions

This chapter describes the identification of promoter fragments that correspond to the cDNAs that were isolated in chapter three. The promoter fragments range in length from 723 bp to 1587 bp, relative to the start site of translation (table 4.1). Previous studies such as Hsu *et al.* (1999), Sunilkumar *et al.* (2002), Yamagata *et al.* (2002), Harmer (2003), Agius *et al.* (2005) and Delaney (2005) have shown that most plant promoters in the range of several hundred base pairs to 1 kb reproduce *in planta* gene expression patterns in experimental

Figure 4.14: Alignment of seven granule-bound starch synthase promoter fragments generated by genomic PCR. The GBSS promoter was generated by PCR with *Taq* polymerase using genomic DNA as template. Seven independently generated PCR products were cloned into pGEMT-Easy, analysed and a consensus sequence determined. In the alignment shown, PCR products #1 and #2 were deemed to represent the *in planta* promoter sequence. The primers used for PCR amplification of the promoter are indicated with red arrows with the included restriction enzyme recognition sites indicated in red.




720
740
760
780
800
820
840
860
880
900
920
940
960
980
1000
1020
1040
Gen.Walk :
PCR#1 :
PCR#2 :
PCR#3 :
PCR#4 :
PCR#5 :
PCR#6 :
PCR#7 :
XbaI

systems, indicating that the isolated promoter fragment are likely to contain the CAREs required for high level boll wall expression. The promoters were isolated by PCR using genomic DNA as template and cloned into the promoterless pJK reporter vector.

Promoter	Promoter length (pre-ATG)	Promoter length (pre-TSS)	Promoter length cloned
PRP	1587 bp	1570 bp	1582 bp
CHS	723 bp	646 bp	708 bp
GBSS	1032 bp	918 bp	1032 bp

Table 4.1: Putative promoters for transient expression analysis.
Various lengths of promoter fragments were identified for each of the three cDNAs. The promoter lengths are shown relative to the start site of translation (pre-ATG) and the start site of transcription (pre-TSS). Also shown are the lengths of each promoter that was cloned into the pJK reporter vector.

The following chapter will describe the transient transformations of cotton tissues with reporter constructs containing the PRP, GBSS and CHS promoter fragments, which were identified by Genome Walker® PCRs in this chapter. As well as assessing the patterns of GUS expression for each promoter, computational analysis of the isolated promoters will be described.

Chapter 5: Functional and bioinformatic analysis of boll wall-preferential promoters

5.1: Introduction

The two previous chapters describe the identification of cDNAs that correspond to mRNA transcripts that are more abundant in the boll wall than leaf, and the subsequent isolation of the corresponding promoter regions. To test the expression driven by the putative promoters, reporter constructs were generated and transient transformations of various cotton tissues performed. In addition to functional promoter analysis, bioinformatic methods were utilised to identify CAREs (*cis*-acting regulatory elements) within the promoters. The functional and computational analysis of the promoter fragments is described in this chapter.

5.2: Transient transformation of cotton tissues

5.2.1: Background

Methods of transient gene expression in many plant species are well established as fast alternatives to the generation of stably transformed plants. Transient gene expression is frequently achieved by the bombardment of DNA coated-microprojectiles into plant tissues. In cotton, microprojectile bombardment has been used to analyse reporter expression patterns that are driven by promoters (Harmer *et al.*, 2002, Kim *et al.*, 2002, Delaney, 2005).

5.2.2: Optimisation and controls

As described in chapter two, a custom-made particle injection gun based on Vain *et al.* (1993) was used for delivery of DNA-coated gold particles into various cotton tissues. There are several variable settings for the gun that can be adjusted depending on the characteristics of the tissue to be transformed. The helium output pressure can be altered in order to propel the DNA-coated gold particle into the target tissue at different speeds. The distance at which the target tissue is located relative to the point of particle injection can also be adjusted between 60 mm and 285 mm. Higher helium output pressures and closer proximity to the injection point result in higher microprojectile velocities and greater penetration into the plant tissue. However, with increased penetration, cells suffer more damage, which can lead to increased cell death. Therefore, settings were selected that allow sufficient tissue penetration, but minimal cell death.

To investigate the expression patterns characteristic of the isolated promoters, a developmental series of boll wall tissue (3, 5, 10, 20, 25 and 30 DPA), together with leaf, bract, calyx, fibre and petal tissues were bombarded with promoter-reporter constructs.

Conditions for microprojectile bombardment of cotton fibre and leaf tissues have been previously described by Harmer *et al*., (2002) and Delaney (2005) and were replicated in this study. These conditions were a helium output pressure of 90 psi and a bombardment distance of 110 mm. Three or four DPA cotton fibres were used for bombardment due to their previously established high levels of expression of transformed genes (Harmer, 2003) and young leaves, prior to senescence, were selected on the same basis. The optimal conditions for other tissues were determined by bombardment with the *35S::GUS* construct (figure 5.2).

It was necessary to use thin, flat sections of boll wall tissue (as described in chapter two) for optimal transformation rates, which may be due to an increase in staining efficiency in these thinner sections. The transformation efficiency of boll wall tissue was further improved by increasing the helium injection pressure to 120 psi (data not shown). It is likely that the boll wall tissue required higher microprojectile velocities compared to other tissues because the boll wall is covered by a waxy layer and has a think outer layer of cortical cells that are tightly packed with little or no intercellular space (Bondada *et al.*, 1994). Reporter expression was seen in 3, 5, 10, 20 and 30 DPA boll wall timepoints (figure 5.1). Relatively weak expression was seen at 3 DPA and GUS activity was not seen in tissue younger than 3 DPA. High levels of GUS activity were seen between 5 and 30 DPA.

Several timepoints of bract and calyx tissue were tested for transformation efficiencies with 1 DPA bract and 13 DPA calyx giving the best results (data not shown). Using an output pressure of 90 psi, bracts were bombarded at a distance of 110 mm and calyces at a distance of 185 mm (figure 5.2). The bombardment conditions for petals were also optimised with a bombardment distance of 135 mm and a helium output pressure of 90 psi used for 0 DPA petals (figure 5.2).

During optimisation with the *35S::GUS* construct, a range of sizes and intensities of GUS-expressing foci were seen for all tissues, indicating varying levels of the GUS reporter protein in transformed cells. This range of expression may be due to: transformation of varying numbers of neighbouring cells, transformation of cells with varying numbers of the reporter construct or cell damage and subsequent 'leakage' of GUS protein into neighbouring cells. Expression of the reporter gene in the different boll wall timepoints was similar, both in number of foci and their intensity, with weaker transformation observed at 3 DPA (figure 5.1). The technical efficiency of the transformation procedure in non-boll wall tissues was high, producing a range of sizes and intensities of GUS-expressing foci (figure 5.2). In addition, biolistics and subsequent histochemical staining of cotton tissues resulted in no

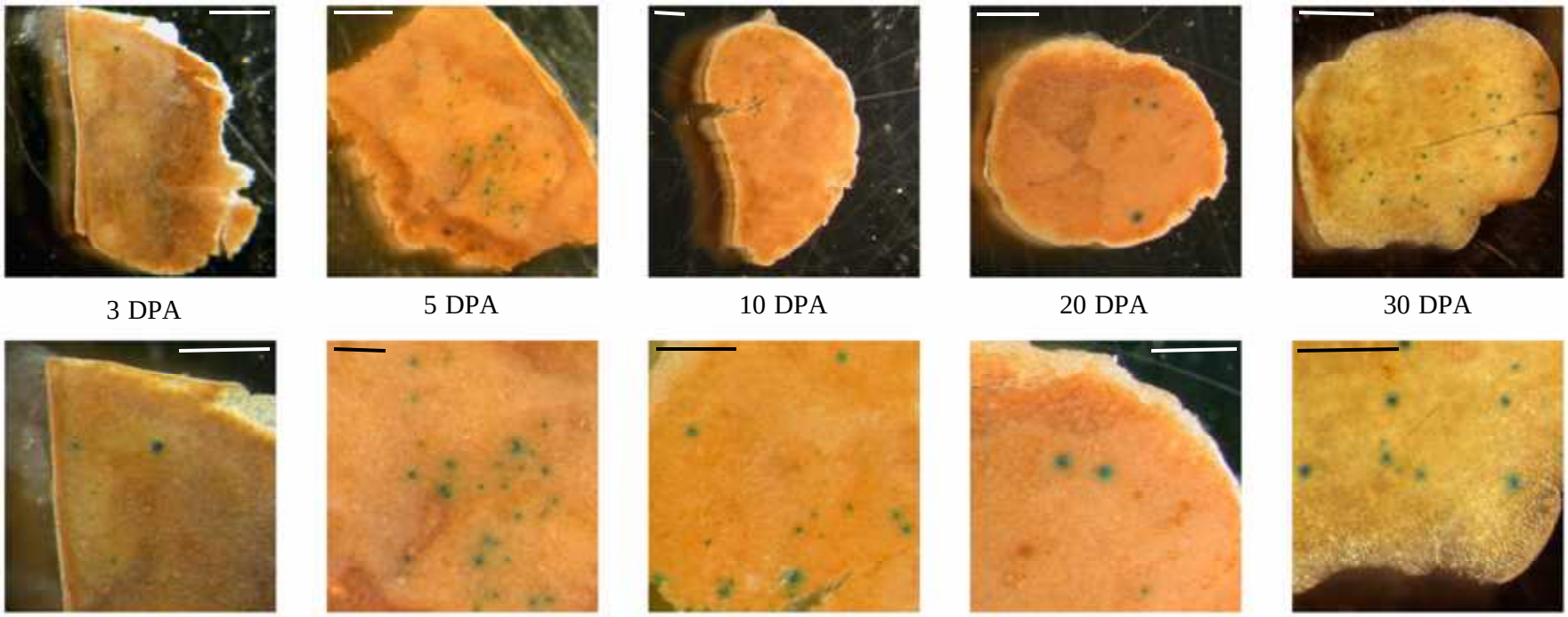


Figure 5.1: Particle bombardment of boll wall tissues with *35S::GUS*.
Boll wall timepoints, as indicated, were transiently transformed with the *35S::GUS* reporter construct. Two images of the transformation results for each timepoint are shown, with images in the bottom panels at a higher magnification than the corresponding images in the top panel. Scale bars represent 2 mm in the top panels and 1 mm in the bottom panels.

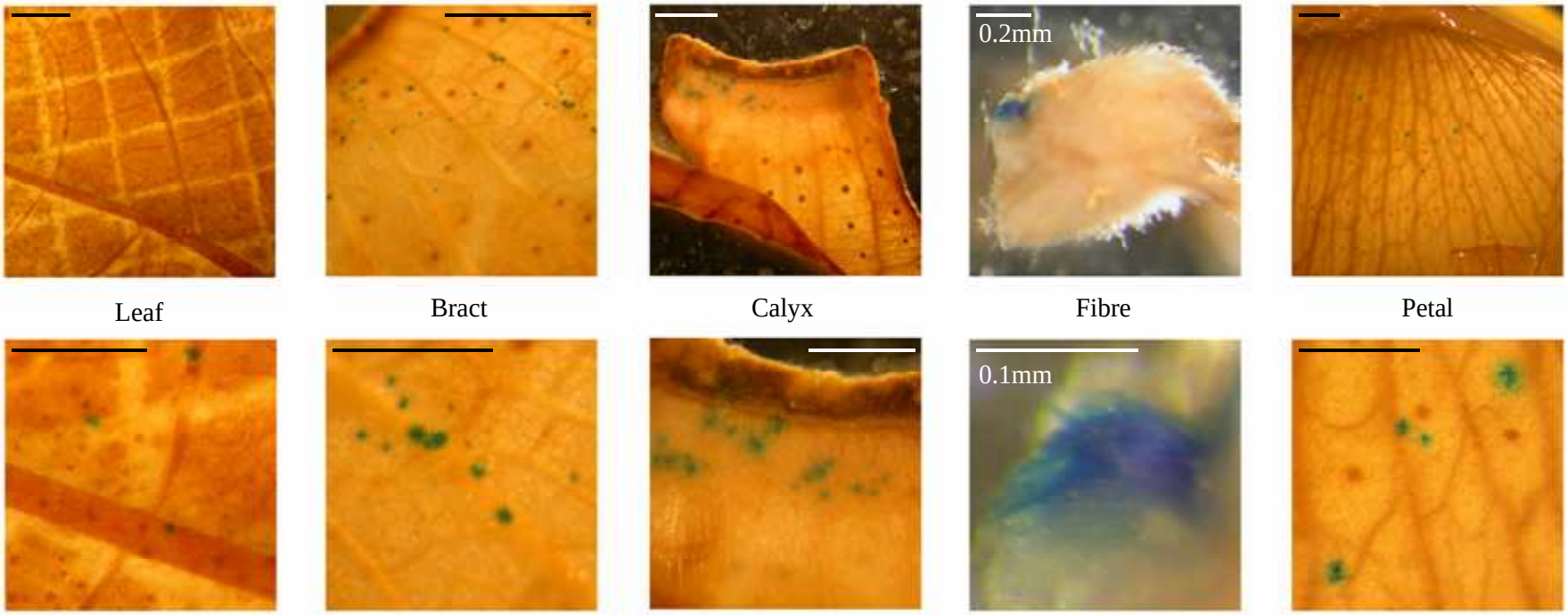


Figure 5.2: Particle bombardment of leaf, bract, calyx, fibre and petal tissues with *35S::GUS*.
Tissues, as indicated, were transiently transformed with the *35S::GUS* reporter construct. Two images of the transformation results for each timepoint are shown, with images in the bottom panels at a higher magnification than the corresponding images in the top panel. Scale bars represent 2 mm in the top panels and 1 mm in the bottom panels, except where otherwise indicated.

GUS-expressing foci in negative controls (empty vector), indicating that the procedure does not produce false positive results.

A major problem with the method of microprojectile bombardment is the lack of an appropriate positive control for each bombardment. To overcome this, each tissue was bombarded in at least two independent transformation experiments, with those tissues that gave a negative result bombarded in at least three independent transformation experiments. Each experiment usually comprised 8 or 16 individual bombardments, with each tissue type bombarded two or four times. Therefore, a tissue that gave a negative transformation result was bombarded a total of 6 to 12 times in three independent experiments. Usually, where negative results were observed, other tissues that were bombarded concurrently showed GUS-expressing foci. Thus, while there was no positive control for each individual bombardment, the methods described were used to confirm that negative expression results are likely to be a real reflection of promoter activity.

5.2.3: Biolistic analysis of cotton tissues with *PRP::GUS*

Biolistic analysis of boll wall tissues with the *PRP::GUS* construct resulted in GUS-expressing foci at 5, 10, 20 and 25 DPA and no expression at 3 and 30 DPA (figure 5.3). In general, between one and five foci of varying sizes were observed. While, it is not possible to quantify the number or strength of GUS-expressing foci, generally they are smaller and fewer in number in the tissues bombarded by *PRP::GUS* than in those bombarded by *35S::GUS.* This is to be expected since the *35S* promoter has been previously described as driving very high levels of expression (Odell *et al.*, 1985, Jefferson *et al.*, 1987). The expression pattern of the reporter gene in the boll wall is similar to the transcription pattern of the PRP transcript, as revealed by Northern analysis (figure 3.7). PRP mRNA was detected between 0 and 25 DPA, with high levels of transcript between 5 and 15 DPA.

Bombardment of leaf, bract, calyx, fibre and petal tissues with the *PRP::GUS* construct resulted in reporter expression only in fibres. From three independent experiments performed on fibres, only one transformed fibre cell produced enough GUS protein to be detectable. Northern analysis did not detect the presence of PRP mRNA in fibres (figure 3.7) but it is possible that a low level of transcript, below the sensitivity limit of Northern analysis, is present. The spatial and temporal patterns of GUS expression driven by the isolated PRP promoter are similar to the expression pattern of the mRNA detected by Northern analysis, which suggests that the isolated promoter fragment contains the elements required for the specificity of PRP gene expression *in planta.*

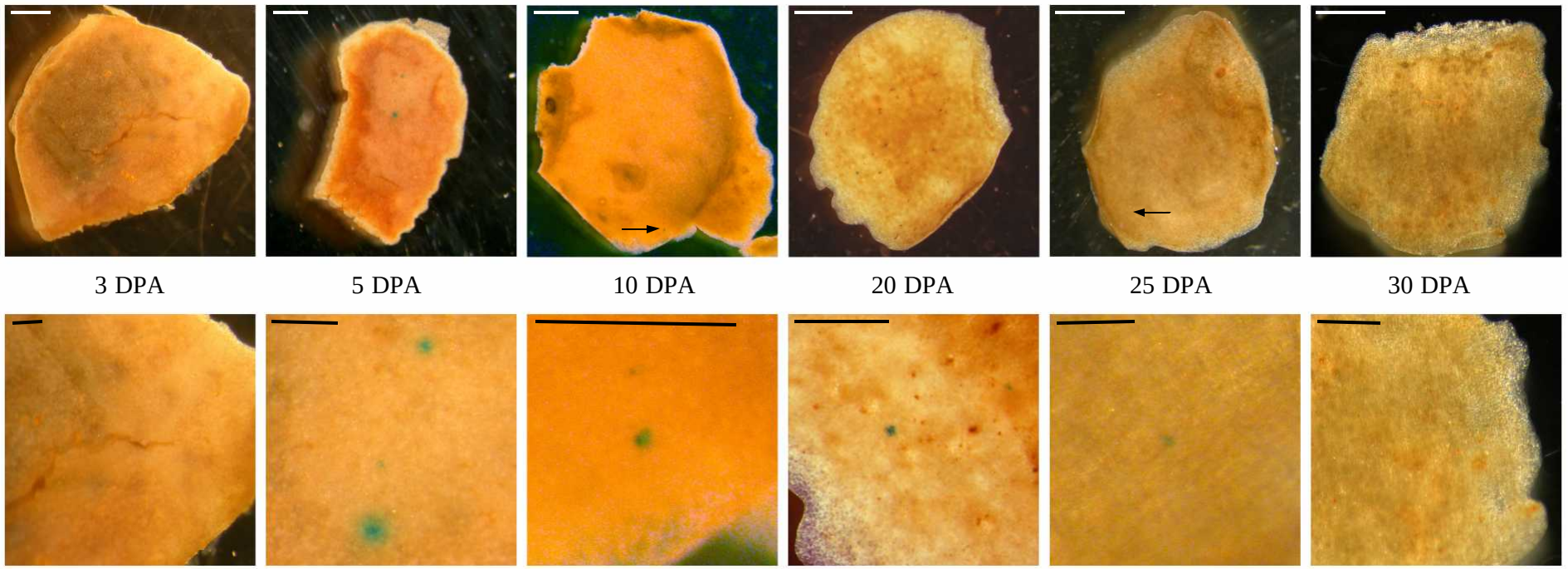


Figure 5.3: Particle bombardment of boll wall tissues with *PRP::GUS*.
Boll wall timepoints, as indicated, were transiently transformed with the *PRP::GUS* reporter construct. Two images of the transformation results for each timepoint are shown, with images in the bottom panels at a higher magnification than the corresponding images in the top panel. Arrows in the top panels indicate GUS-expressing foci that are shown in the corresponding bottom panels. Scale bars represent 2 mm in the top panels and 1 mm in the bottom panels.

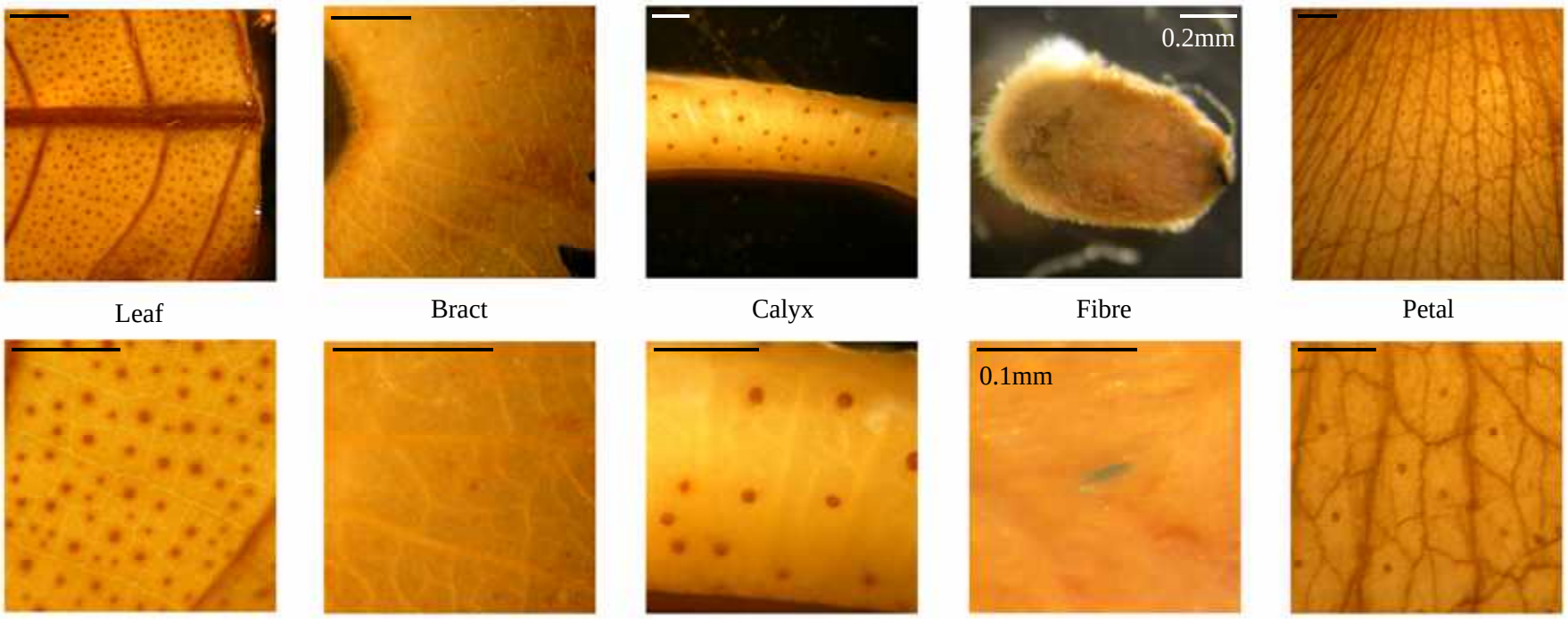


Figure 5.4: Particle bombardment of leaf, bract, calyx, fibre and petal tissues with *PRP::GUS*.
Tissues, as indicated, were transiently transformed with the *PRP::GUS* reporter construct. Two images of the transformation results for each timepoint are shown, with images in the bottom panels at a higher magnification than the corresponding images in the top panel. Scale bars represent 2 mm in the top panels and 1 mm in the bottom panels, except where otherwise indicated.

5.2.4: Biolistic analysis of cotton tissues with *CHS::GUS*

Transient transformation of boll wall tissue with the *CHS::GUS* construct revealed GUS protein foci at 3, 5, 10, 20 and 25 DPA (figure 5.5). Temporal expression of the reporter in the boll wall was similar when driven by either the CHS or PRP promoters except for positive expression at 3 DPA with the *CHS::GUS* construct. As was the case for *PRP::GUS*, generally between one and five GUS protein foci of varying sizes were observed in sections of boll wall positively transformed with *CHS::GUS*. The temporal expression pattern is similar to the transcription pattern of CHS mRNA (figure 3.7), which was detected between 0 and 25 DPA, with highest abundance between 0 and 15 DPA.

In addition to the boll wall, the *CHS::GUS* construct also drove expression of the reporter gene in fibre and petal tissues (figure 5.6). No reporter gene expression was observed for leaf, bract and calyx tissues (figure 5.6). As can be seen in figure 3.7, the CHS mRNA was detected in fibre and petal tissues, which corresponds to the expression of the reporter gene driven by the CHS promoter. Thus, as is the case for the PRP promoter, reporter expression correlates well with the transcription pattern of the mRNA, suggesting that the isolated promoter fragment contains the elements required for the specificity of CHS gene expression *in planta*. However, a relatively low rate of transformation was observed in petals, with one focus of GUS-expression observed from three independent transformation experiments. This is in contrast to the high abundance of the CHS transcript observed by Northern analysis.

5.2.5: Biolistic analysis of cotton tissues with *GBSS::GUS*

No GUS-expressing foci were observed in any tissues bombarded with the *GBSS::GUS* construct (figures 5.7 and 5.8). As explained in section 5.2.2, negative bombardment experiments were repeated, with a total of three independent experiments performed with between 6 and 12 individual bombardments of each tissue. The repeated failure of the *GBSS::GUS* construct to yield any GUS-expressing foci combined with the positive results obtained with the *PRP::GUS* and *CHS::GUS* constructs suggests that the isolated GBSS promoter fragment is incapable of driving reporter expression. This result is surprising as the GBSS promoter fragment (1,032 bp) is longer than the CHS promoter fragment (708 bp), which was sufficient to drive expression of the reporter gene. The abundance of the PRP, CHS and GBSS mRNA transcripts were also similar when assessed by Northern analysis (figure 3.7), indicating that the corresponding promoters are of approximately equivalent strength.

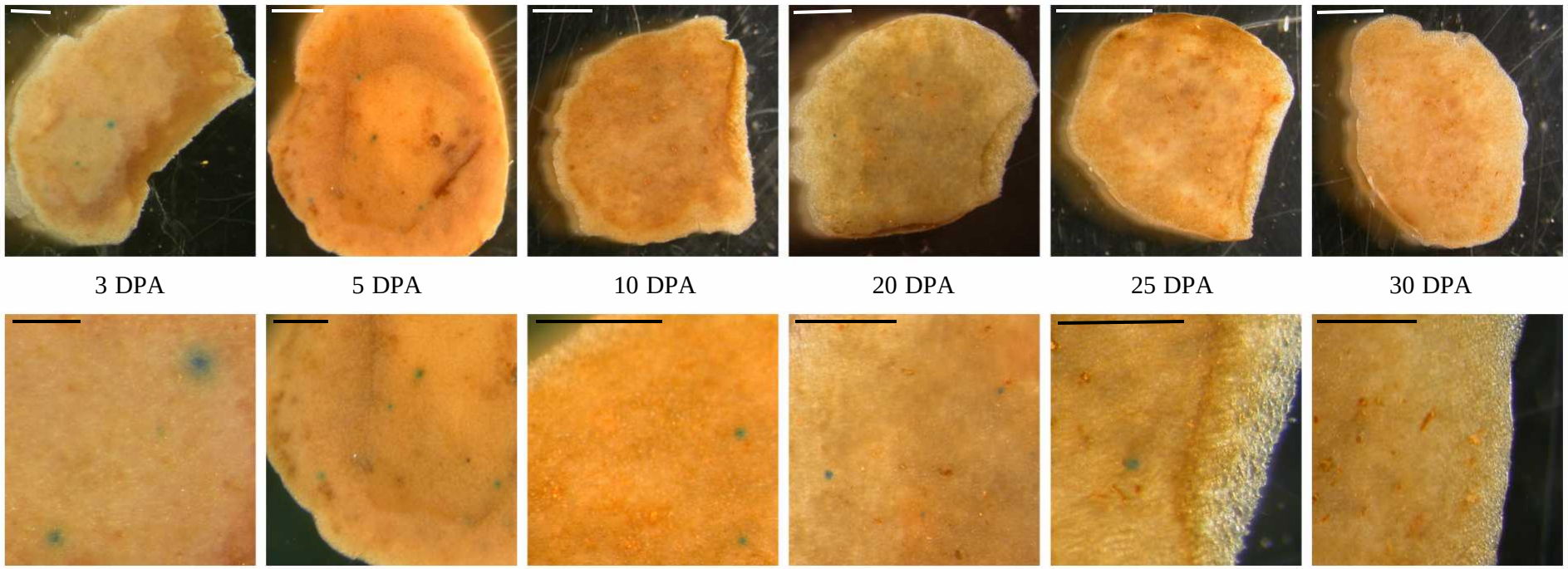


Figure 5.5: Particle bombardment of boll wall tissues with *CHS::GUS*.
Boll wall timepoints, as indicated, were transiently transformed with the *CHS::GUS* reporter construct. Two images of the transformation results for each timepoint are shown, with images in the bottom panels at a higher magnification than the corresponding images in the top panel. Scale bars represent 2 mm in the top panels and 1 mm in the bottom panels.

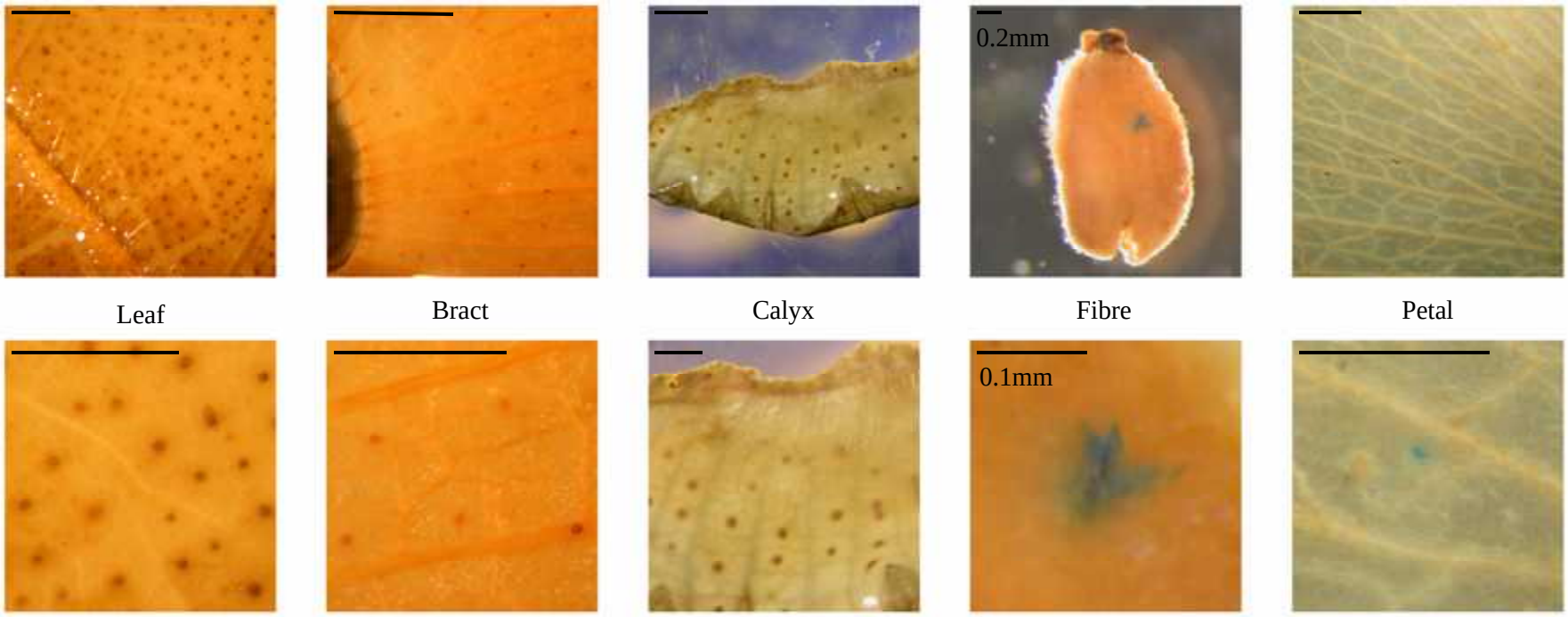


Figure 5.6: Particle bombardment of leaf, bract, calyx, fibre and petal tissues with *CHS::GUS*.
Tissues, as indicated, were transiently transformed with the *CHS::GUS* reporter construct. Two images of the transformation results for each timepoint are shown, with images in the bottom panels at a higher magnification than the corresponding images in the top panel. Scale bars represent 2 mm in the top panels and 1 mm in the bottom panels, except where otherwise indicated.

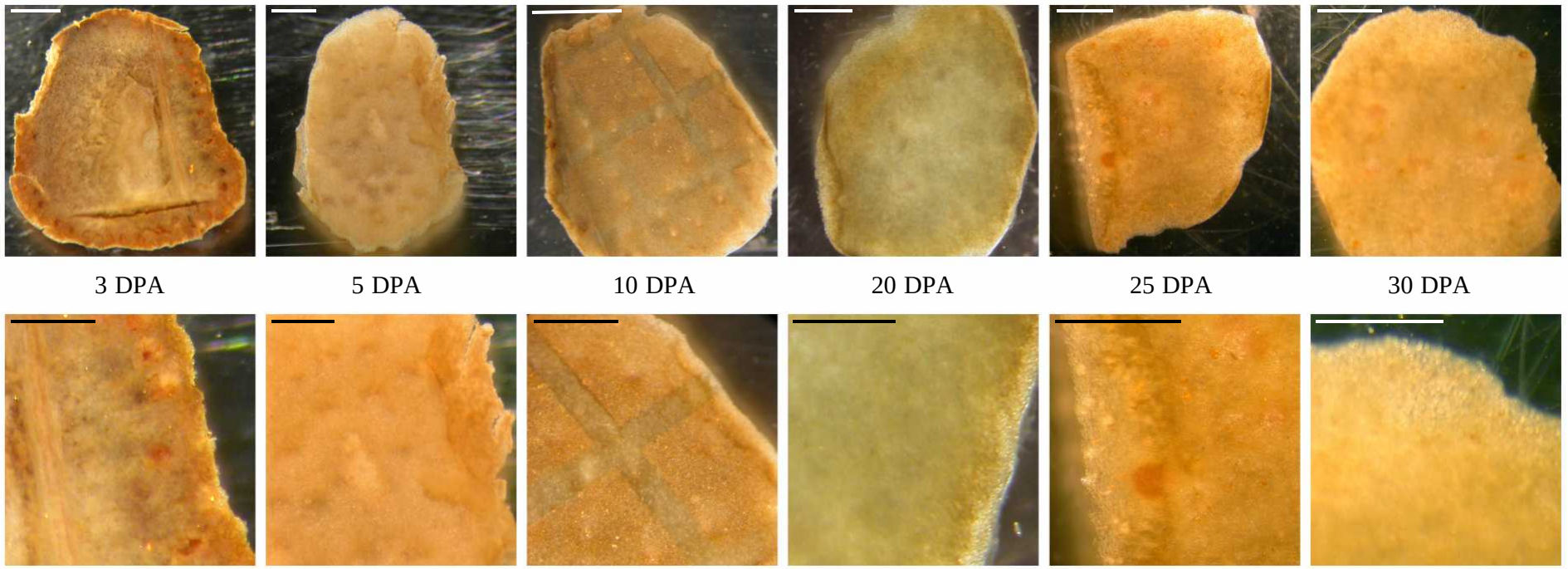


Figure 5.7: Particle bombardment of boll wall tissues with *GBSS::GUS*.
Boll wall timepoints, as indicated, were transiently transformed with the *GBSS::GUS* reporter construct. Two images of the transformation results for each timepoint are shown, with images in the bottom panels at a higher magnification than the corresponding images in the top panel. Scale bars represent 2 mm in the top panels and 1 mm in the bottom panels.

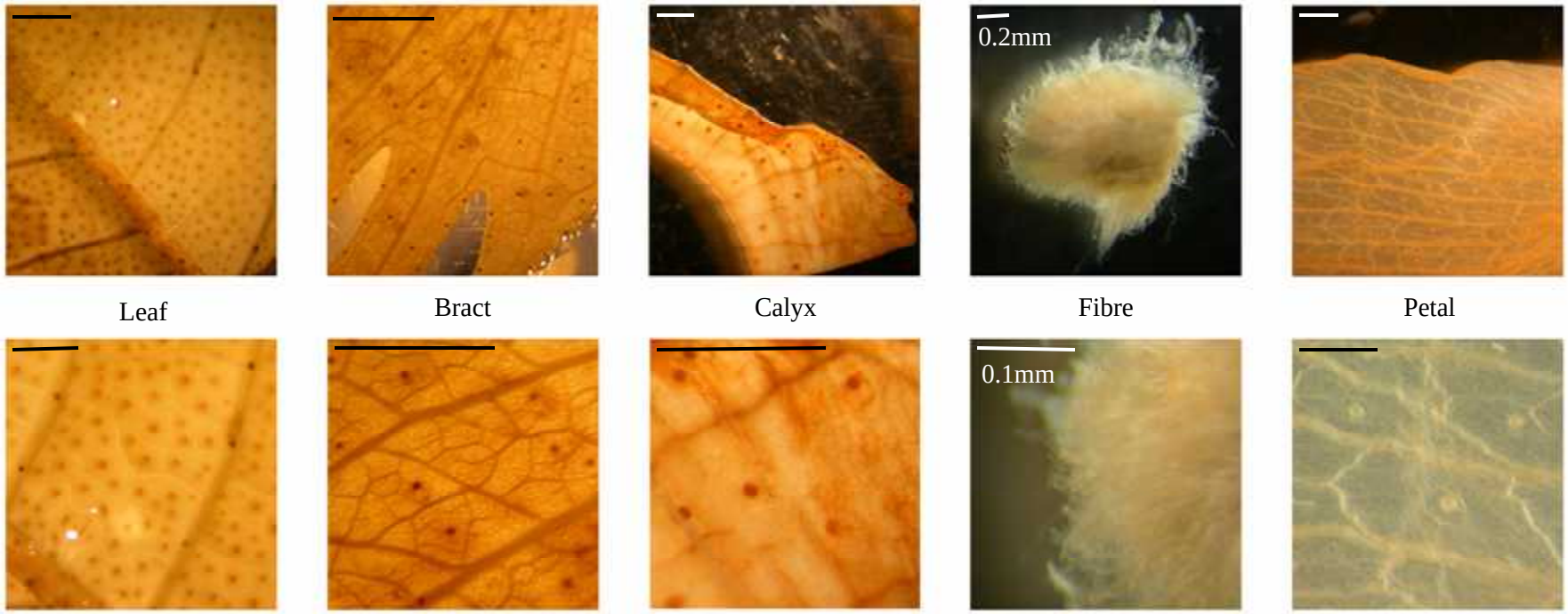


Figure 5.8: Particle bombardment of leaf, bract, calyx, fibre and petal tissues with *GBSS::GUS*.
Tissues, as indicated, were transiently transformed with the *GBSS::GUS* reporter construct. Two images of the transformation results for each timepoint are shown, with images in the bottom panels at a higher magnification than the corresponding images in the top panel. Scale bars represent 2 mm in the top panels and 1 mm in the bottom panels, except where otherwise indicated.

5.2.6: Transient expression discussion and summary

Conditions were optimised for the bombardment of boll wall (3, 5, 10, 20 and 30 DPA), leaf, bract, calyx, fibre and petal tissues. Bombardment of two of the constructs, *PRP::GUS* and *CHS::GUS*, resulted in positive expression while the third construct, GBSS, yielded no expression. The expression pattern of GUS, driven by the PRP and CHS promoter fragments correlates well with their transcript patterns as revealed by Northern analysis (figure 3.7). This similarity indicates that both of these promoter fragments contain the necessary elements for the boll wall-preferential gene transcription that is seen *in planta*.

The lack of positive expression in any tissues bombarded with the *GBSS::GUS* construct is surprising given the high levels of the transcript detected by Northern analysis and the length of promoter used in the transient assay. Possible explanations include the isolation of a promoter fragment that does not contain all of the CAREs required for gene expression, or a technical problem with the identification and isolation of the promoter fragment. The explanation that the promoter elements required for gene expression lie outside of the isolated region seems unlikely as elements for transcription are usually located within the first several hundred base pairs upstream of the transcription start site (Tyagi, 2001). Therefore, a technical problem such as the incorrect identification or isolation of the promoter region corresponding to the GBSS mRNA transcript is the most likely explanation. Such an error may have occurred during the amplification of the promoter by PCR, with a region of similar sequence mistakenly amplified. Alternatively, mutations may have been introduced into the amplified promoter, resulting in the loss of important promoter elements.

5.3: Bioinformatic analysis of boll wall-preferential promoters

5.3.1: Background

Databases that contain experimentally verified CAREs have been assembled, such as PLACE (a database of plant *cis*-acting regulatory DNA elements) (Prestridge, 1991, Higo *et al.*, 1999), and can be utilised to identify potential CAREs within promoters. However, the usefulness of these databases is limited because of the small number of verified CAREs, with many being slight variations of other related elements. Furthermore, many of the promoter motifs contained within the databases are relatively short, with recognition sites of four or five nucleotides, which means that the motifs will be found by chance every 256 or 1024 nucleotides, respectively. Despite these limitations, CARE database analysis can be a useful tool for identifying potential transcription factor binding sites, which can then be experimentally verified. Release 27.0 (May 31, 2006) of the PLACE database

(http://www.dna.affrc.go.jp/PLACE), containing 453 experimentally verified promoter motifs, was used for the promoter analysis described here.

In addition to the identification of previously characterised CAREs, programs have also been designed to identify novel CAREs by comparing promoter sequences of co-regulated genes to identify short regions of homology that are statistically over-represented within promoter sequences. The accuracy of 13 of these programs was assessed by Tompa *et al*. (2005). The authors generated artificial sets of promoters, which shared previously verified CAREs, to test whether the programs could identify the motifs. The programs performed poorly with only a weak correlation between the location of motifs identified by the programs and the location of the known motif. Given the poor performance of such programs, a simple dot plot analysis (Maizel and Lenk, 1981) was used in preference to identify potential novel CAREs in the boll wall promoter set. The promoter sequences were compared with each other using the COMPARE program with a window size of seven nucleotides. This program compares short stretches of sequence between two sequences, with the output file interpreted by the DOTPLOT program, which plots regions of homology on a graphical output. This method allows for the identification of short sequences that would not be identified using standard computational methods such as EclustlW (Thompson *et al.*, 1994).

5.3.2: PLACE analysis of the proline rich protein and chalcone synthase promoters

The promoter regions that were identified by Genome Walker® PCR (up to and including the ATG start site of translation initiation) were compared to verified CAREs in the PLACE promoter database. Analysis of the PRP and CHS promoter sequences resulted in the identification of numerous putative CAREs with 342 identified in the PRP promoter (table 5.1 and appendix 2A) and 168 in the CHS promoter (Table 5.1 and appendix 2B). There was a high level of repetition within the sets of CAREs identified for each promoter, with only 84 and 54 CARE motifs identified as unique to the CHS and PRP promoters, respectively. Putative TATA and CAAT boxes were identified in both promoter sequences that correspond to the consensus sequences for the two motifs (Joshi, 1987b) (figures 5.9 and 5.10).

G-box and H-box-like motifs were identified in the CHS promoter, positioned at –87 to –82 and –74 to –69, relative to the presumed start site of transcription (figure 5.10). CHS promoters from many species contain G-box and H-box motifs that have been shown to be important for CHS expression, specifically in tomato fruit (Baum *et al.*, 1997), tobacco flowers (Faktor *et al.*, 1996, Faktor *et al.*, 1997) and in response to: elicitors in bean (Lawton *et al.*, 1990), light in mustard (Kaiser and Batschauer, 1995) and *p*-coumaric acid in alfalfa

CARE identity	PRP	CHS	GBSS
-10PEHVPSBD	+	+	+
-300ELEMENT	+		+
AACACOREOSGLUB1	+	+	
ABRELATERD1		+	+
ABREZMRAB28			+
ACGTABOX	+		
ACGTABREMOTIFA2OSEM			+
ACGTATERD1	+	+	+
ACGTTBOX			+
AGMOTIFNTMYB2			+
AMMORESIIUDCRNIA1	+		
AMYBOX1	+		
ANAERO1CONSENSUS	+	+	
ANAERO3CONSENSUS	+		+
ARE1			+
ARFAT	+		
ARR1AT	+	+	+
ASF1MOTIFCAMV	+	+	+
BOXIINTPATPB	+		+
BOXIIPCCHS			+
BOXLCOREDCPAL			+
CAATBOX1	+	+	+
CACGCAATGMGH3	+		
CACGTGMOTIF		+	+
CACTFTPPCA1	+	+	+
CAREOSREP1	+	+	+
CARGCW8GAT31	+		
CATATGGMSAUR	+		
CCAATBOX1	+		+
CACGTGMOTIF		+	
CPBCSPOR		+	
CURECORECR	+	+	+
DOFCOREZM111	+	+	+
DPBFCOREDCDC3	+	+	
EBOXBNNAPA	+	+	+
EECCRCAH1	+		
ELRECOREPCRP1	+		
ERELEE4	+		+
GAREAT	+		
GATABOX	+	+	+
GT1CONSENSUS	+	+	+
GT1CORE	+		
GT1GMSCAM4	+	+	+
GT1MOTIFPSRBCS	+		
GTGANTG10	+	+	+
HEXMOTIFTAH3H4		+	
IBOX883	+		
IBOXCORE	+		+
INRNTPSADB	+		+

INTRONLOWER			+
LECPLEACS2		+	
LTRE1HVBLT49	+	+	
LTRECOREATCOR15	+		
MARTBOX			+
MRNA3ENDTAH3	+		
MYB1AT	+		+
MYB2AT			+
MYB2CONSENSUSAT		+	+
MYBCORE	+	+	+
MYBGAHV	+		
MYBPLANT			+
MYBPZM	+		+
MYBST1	+		+
MYCATERD1	+	+	
MYCATRD22	+	+	
MYCCONSENSUSAT	+	+	+
NODCON1GM	+	+	+
NODCON2GM	+	+	+
NTBBF1ARROLB	+		
OSE1ROOTNODULE	+	+	+
OSE2ROOTNODULE	+	+	
P1BS	+		
POLLEN1LELAT52	+	+	+
PREATPRODH	+		+
PROLAMINBOXOSGLUB1	+		
PYRIMIDINEBOXOSRAMY1A	+	+	+
RAV1AAT	+	+	+
RBCSCONSENSUS			+
REALPHALGLHCB21	+	+	+
ROOTMOTIFTAPOX1	+	+	+
RYREPEATBNNAPA	+		
RYREPEATGMGY2	+		
RYREPEATLEGUMINBOX	+		
S1FBOXSORPS1L21	+	+	+
SEBFCONSSTPR10A	+		
SEF1MOTIF	+	+	+
SEF3MOTIFGM			+
SEF4MOTIFGM7S	+	+	+
SORLIP1AT	+		+
SORLIP2AT			+
SORLREP2AT			+
SREATMSD	+		
SV40COREENHAN			+
T/GBOXATPIN2		+	+
TAAAGSTKST1	+	+	+
TBOXATGAPB	+		
TGACGTVMAMY		+	
TRANSINITDICOTS	+		
TRANSINITMONOCOTS	+		

UP2ATMSD	+		
UPRMOTIFIIAT			+
WBOXATNPR1	+	+	+
WBOXHVISO1	+	+	+
WBOXNTERF3	+	+	+
WRKY71OS	+	+	+

Table 5.1: Putative CAREs identified in the proline rich protein, chalcone synthase and granule-bound starch synthase promoters by PLACE analysis. The CAREs identified by PLACE analysis in the three promoter fragments are shown above. Elements such as polyadenylation signals and TATA boxes have been removed for this analysis. The presence of a CARE within a promoter is indicated by a '+' in the corresponding cell. CAREs common to all three promoters are indicated in red, CAREs common to the PRP and CHS promoters are indicated in blue.

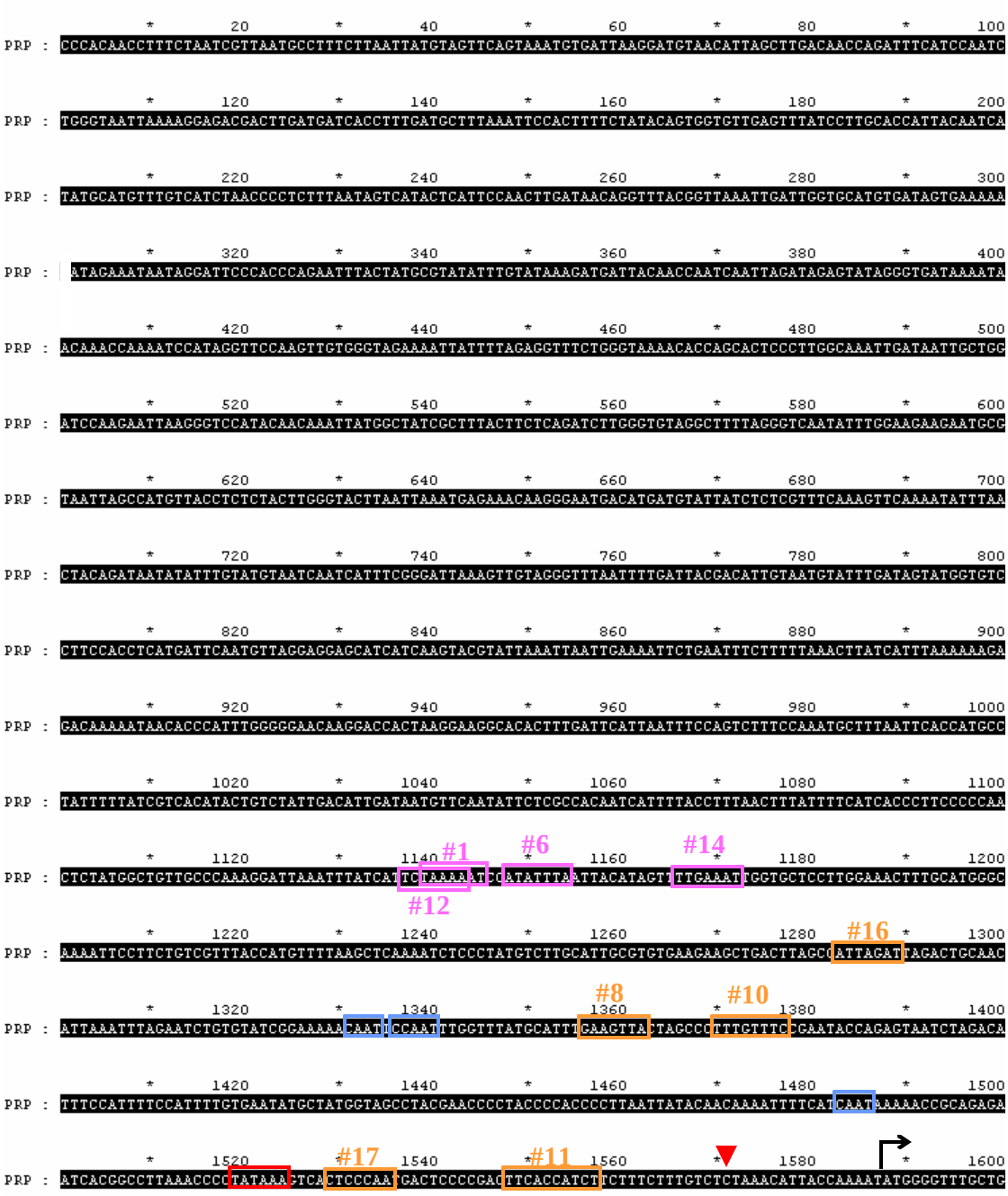


Figure 5.9: Proline rich protein promoter region.
The PRP promoter region and the 5' end of the PRP mRNA transcript are shown above. The putative start sites of translation and transcription are shown with a black arrow and a red triangle, respectively. Potential TATA and CAAT boxes are indicated with red and blue boxes, respectively. Orange boxes indicate potential CAREs identified in the PRP and chalcone synthase promoters that are within 300 bp of the putative transcription start site and pink boxes indicate novel potential CAREs that were present in the PRP, CHS and GBSS promoters.

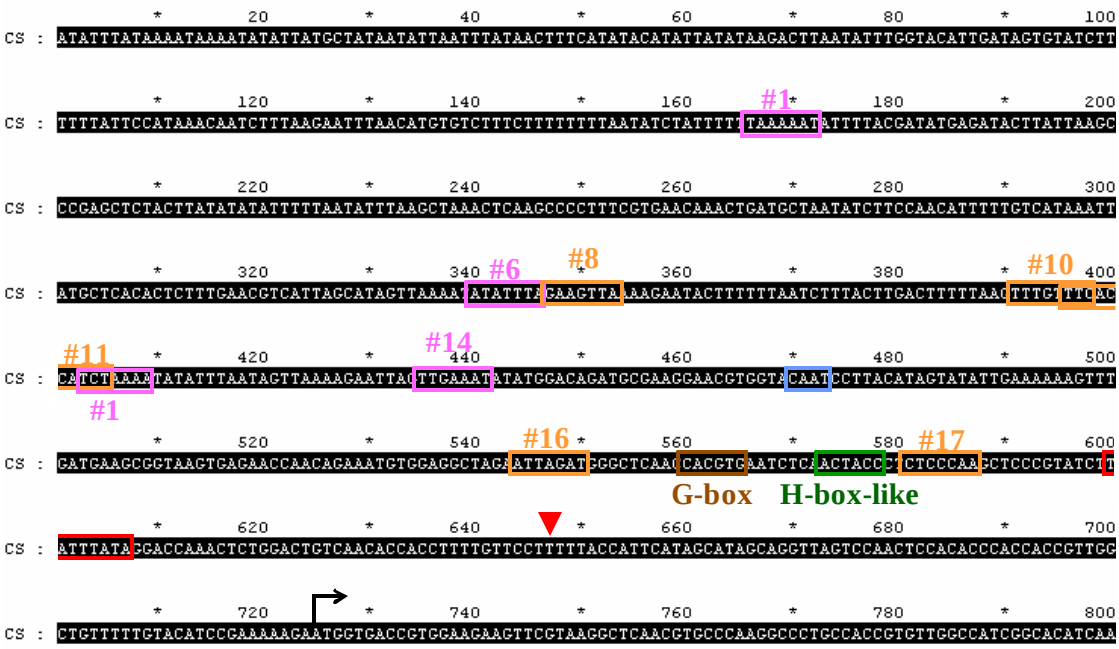


Figure 5.10: Chalcone synthase promoter region.
The CHS promoter region and the 5' end of the CHS mRNA transcript are shown above. The putative start sites of translation and transcription are shown with a black arrow and a red triangle, respectively. Potential TATA and CAAT boxes are indicated with red and blue boxes, respectively. Brown and green boxes show G-box and H-box-like motifs, respectively. Orange boxes indicate potential CAREs identified in the PRP and CHS promoters that are within 300 bp of the putative transcription start site and pink boxes indicate novel potential CAREs that were present in the PRP, CHS and GBSS promoters.

(Loake *et al.*, 1992). A 5'-CACGTG-3' sequence motif was identified as a "CACGTGMOTIF" CARE (table 5.1 and appendix 2B), but is also known as a G-box (Lawton, 1990; Faktor, 1996). The H-box-like motif was identified by comparison of the cotton CHS promoter with other CHS promoters, where G-box and H-box motifs are often found in close proximity (Faktor *et al.*, 1996, Faktor *et al.*, 1997, Sakuta, 2000). The H-box-like motif identified here (5'-ACTACC-3') is similar to H-box motifs (5'-CCTACC-3'), found in many CHS promoters.

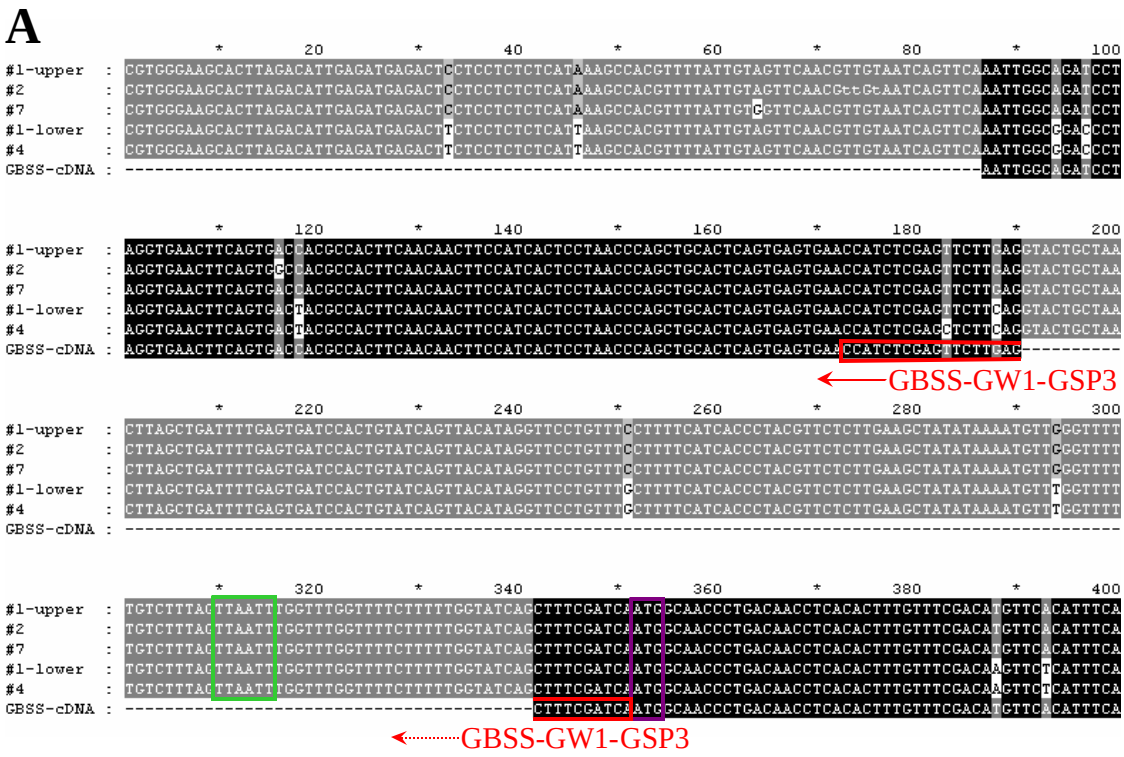
5.3.3: PLACE analysis of the granule-bound starch synthase promoter

The GBSS promoter did not drive reporter expression in any tissues tested. Potential explanations for the lack of expression include the amplification of non-target promoter fragments by Genome Walker® PCR or the introduction of mutations during amplification from genomic DNA. Further analysis of this region revealed the likely cause of lack of reporter expression to be the first of these explanations.

The region of sequence between nucleotides 338 and 441 in figure 4.7 was incorrectly aligned with the GBSS cDNA sequence. This region was previously characterised as non-target Genome Walker® PCR product (section 4.3.2) but is probably an intron (figure 5.11A). The tertiary Genome Walker® PCR primer (GBSS-GW1-GSP3) amplified sequence from the exon upstream of the intron, causing the omission of the intron from the amplified promoter fragment. The putative intron was eventually identified by eye during bioinformatic analysis of the GBSS promoter region, after traditional sequence alignment programs such as EclustlW (Thompson *et al.*, 1994) failed to correctly align the 5' exon.

The 5' (AG/gt) and 3' (ag/CT) intron boundaries match the consensus intron boundary sequences of dicot plants (Simpson and Filipowicz, 1996). A potential branch site (TTAAT) is located 29 nucleotides from the 3' intron splice site, which is similar to the consensus site of CT(G/A)A(C/T). The intron is also 65% adenine and thymine, which is consistent with most plant introns (Simpson and Filipowicz, 1996).

The presence of an intron in the 5' untranslated region of the GBSS promoter is not unusual as similar introns have been identified in GBSS genes of rice (Yu *et al.*, 1992), potato (Fu *et al.*, 1995), maize (Shaw *et al.*, 1994), *Arabidopsis* (Chopra *et al.*, 1992) and barley (Patron *et al.*, 2002). The omission of the intronic sequence from the GBSS promoter fragment that was generated in this study could have resulted in a non-functional GBSS promoter due to the loss of transcriptional enhancer elements that may have been present in



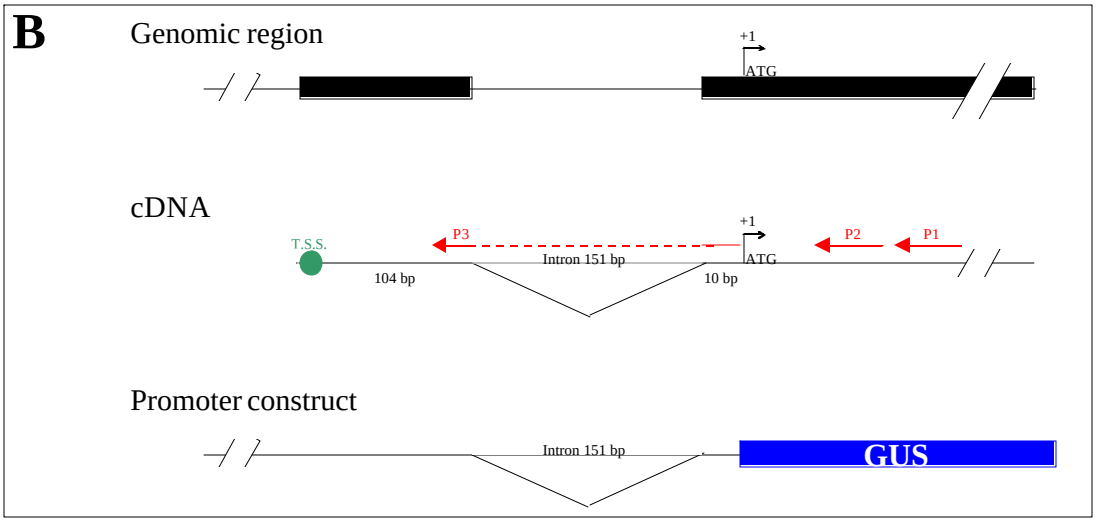


Figure 5.11: Granule-bound starch synthase genomic region.
An alignment of the GBSS cDNA with the Genome Walker® PCR products that were generated for the GBSS promoter is shown above. When compared to figure 3.7, it can be seen that an intron in the 5' UTR was omitted from the promoter sequence of GBSS. The potential binding sites for the tertiary primer used in Genome Walker® PCR are indicated in red, on both sides of the intron. The start site of translation is indicated in purple, and a green box indicates the putative intron branch point. Figure 5.11B shows a schematic diagram of the GBSS promoter and cDNA regions. The genomic region panel shows the known exons (black boxes) and introns (thin black lines). The middle panel shows the cDNA that is generated from the genomic region, with the intron spliced out of the pre-mRNA sequence. This is representative of the cDNA isolated from the original library screen. The primers that were used for Genome Walker® PCR are indicated in red, including P3 (GBSS-GW1-GSP3), which 'spanned' the intron. The presumed start site of transcription is indicated in green. The promoter construct panel shows the promoter fragment that was cloned for transient transformation. The intronic sequence that is part of the promoter region, but not contained within the cDNA, was omitted.

the 5' UTR (Gutierrez-Alcala *et al.*, 2005). Another explanation for the loss of promoter activity is a process known as intron-mediated enhancement of gene expression (IME) (reviewed by Rose and Beliakoff, 2000; Rose, 2004), which occurs in mammals, insects, nematodes and plants (Bhattacharyya *et al.*, 2003). Several studies have compared transcript levels driven by promoter constructs that are identical, except for a 5' UTR intron. These studies have demonstrated up to 1000 fold higher levels of mRNA accumulation in cells transformed with the promoter construct containing the 5' UTR intron compared to cells transformed with the promoter construct lacking the 5' UTR intron (Callis *et al.*, 1987, Maas *et al.*, 1991, Sivamani and Qu, 2006). The mechanism by which the 5' UTR intron affects mRNA accumulation is not known but several studies have suggested that the intron increases transcript stability leading to higher mature mRNA abundance (Johnson *et al.*, 1998). The lack of reporter expression could be explained if the GBSS 5' UTR intron contains important enhancer elements or mediates IME.

Another feature of the GBSS promoter is the lack of a TATA box within the promoter region. The TATA box is a *cis*-acting regulatory region that is found in most gene promoters, which is required for the binding of transcription factors and histones and, in most cases, is essential for the initiation of gene transcription. However, up to 10% of plant genes do not possess a recognisable TATA box (Nakamura, 2002). In TATA-less promoters, pyrimidine-rich initiator (Inr) elements near the transcription start site, have been shown to replace the TATA box and these direct basal transcription rates (Smale and Baltimore, 1989). The consensus sequence for mammalian Inr elements is (T/C)-(T/C)-A-N-(T/A)-(T/C)-(T/C), where adenine at position three is usually the start site of transcription (Smale *et al.*, 1998). In tobacco, an Inr consensus sequence of (T/C)-T-C-A-N-T-(T/C)-(T/C) was established by Nakamura *et al*. (2002). Several sequences similar to these consensus Inr sequences are located near the presumed start site of transcription of the GBSS cDNA, with several more located further upstream (table 5.2).

The entire computationally compiled GBSS promoter sequence, including the 5' UTR intron that was omitted during Genome Walker® PCR, was used for bioinformatic promoter analysis. While this sequence has not been demonstrated to drive expression, this seems most likely given the high GBSS transcript abundance (figure 3.7) and the reporter gene expression driven by the other two promoters fragments of similar length. Therefore, conclusions drawn from the analysis of the GBSS promoter require experimental confirmation.

When the presumptive GBSS promoter is computationally compiled, a total of 1138 bp of sequence upstream of the start site of translation is present. By comparison with the PLACE database, 258 potential CAREs were identified (table 5.1 and appendix 2C). 35 of these were located within the putative intron, including several potential MYB transcription binding sites (appendix 2C).

Mammalian consensus sequence:	-	(C/T)	(C/T)	A	N	(A/T)	(C/T)	(C/T)
Tobacco consensus sequence:	(C/T)	T	C	A	N	T	(C/T)	(C/T)
+8	G	G	C	A	G	A	T	C
+1	T	C	A	A	A	T	T	G
-1	T	T	C	A	A	A	T	T
-6	A	T	C	A	G	T	T	C
-23	T	T	C	A	A	C	G	T
-28	T	T	T	A	T	T	G	T
-34	C	A	T	A	A	A	G	C

Table 5.2: Putative GBSS promoter Initiator (Inr) elements.
Several potential Inr elements identified near the putative start site of transcription of the GBSS cDNA are similar to the consensus Inr elements from mammals and tobacco (Nakamura *et al.*, 2002). The indicated position of the Inr element is the location of the conserved adenine relative to the presumed start site of transcription.

5.3.4: PLACE analysis summary

The CAREs that were identified in each of the three promoters are shown in table 5.1 with those common to the PRP and CHS promoters shown in table 5.3. 38 CARE motifs are located within both the PRP and CHS promoters, with 31 of these also contained within the GBSS promoter. These common CAREs may have roles in response to light, copper, oxygen, salt, pathogens, low temperature, water stress, sugar, wounding and plant growth hormones. CAREs were also identified that are required for tissue-specific expression of their corresponding genes in seeds (*Brassica napus*), leaf mesophyll (*Flaveria trinervia*), pollen (*Nicotiana tabacum* and *Lycopersicon esculentum*) and root (*Nicotiana tabacum*). Potential binding sites for MYB, WRKY and Dof transcription factors, which have important roles in plant development, were also identified. Despite the identification of many common CAREs, there is no evidence to suggest the involvement of any particular CAREs in boll wall-preferential gene transcription.

5.3.5: Dot plot analysis

To identify novel promoter motifs that may be necessary for boll wall gene transcription, dot plot analysis was used to search for short regions of homology in the

CAREs	ID #	Motif	Function (from PLACE database)
-10PEHVPSBD	S000392	TATTCT	Light responsive promoter element identified in the barley chloroplast *psbD* gene promoter.
AACACOREOSGLUB1	S000353	AACAAAC	Essential for the seed specific expression of *GluB-1* in rice.
ACGTATERD1	S000415	ACGT	Required for etiolation-induced expression of *erd1* in *Arabidopsis*.
ANAERO1CONSENSUS	S000477	AAACAAA	Anaerobic response element from maize and *Arabidopsis*.
ARR1AT	S000454	NGATT	*Arabidopsis* ARR1 (*Arabidopsis* response regulator 1) -binding element.
ASF1MOTIFCAMV	S000024	TGACG	ASF1-binding site in CaMV 35S promoters, which may be responsive to auxin, salicylic acid or light.
CACTFTPPCA1	S000449	YACT	Responsible for the mesophyll-specific expression of *ppcA1* in *Flaveria trinervia*.
CAREOSREP1	S000421	CAACTC	Rice gibberellin-response element identified in *Rep1* gene promoter.
CURECORECR	S000493	GTAC	Copper-response element identified in the *Cyc6* and *Cpx1* gene promoters in *Chlamydomonas*. May also be involved in oxygen-response of these genes.
DOFCOREZM	S000265	AAAG	Required for binding of Dof (DNA-binding with one zinc finger) proteins in maize.
DPBFCOREDCDC3	S000292	ACACNNG	Binding site for the bZIP transcription factors, DPBF-1 and DPBF-2 in the carrot *Dc3* gene promoter. May be involved in abscisic acid induction and embryo specificity of the *DcS* promoter.
EBOXBNNAPA	S000144	CANNTG	Necessary for seed expression of *napA* seed storage protein in *Brassica napus*.
GATABOX	S000039	GATA	Promoter motif identified from several species, which binds ASF2 (activating sequence factor 2). Required for light-regulated and tissue-specific gene expression.
GT1CONSENSUS	S000198	GRWAAW	Light-responsive motif found in many light-regulated genes from many species.
GT1GMSCAM4	S000453	GAAAAA	Identified in the *SCaM-4* gene promoter from soybean and is pathogen- and salt-responsive.
GTGANTG10	S000378	GTGA	Element required for pollen-specific expression of the *g10* gene in tobacco.
LTRE1HVBLT49	S000250	CCGAAA	A low temperature responsive element in the *blt4.9* gene promoter of barley.
MYBCORE	S000176	CNGTTR	Binding site for MYB proteins such as ATMYB1 and ATMYB2 from *Arabidopsis* and MYB.Ph3 from petunia, which are involved in water stress response and flavonoid biosynthesis, respectively.
MYCATERD1	S000413	CATGTG	Recognition site necessary for expression of *erd1* in *Arabidopsis* in response to dehydration.
MYCATRD22	S000174	CACATG	Recognition site necessary for expression of *rd22* in *Arabidopsis* in response to dehydration.
MYCCONSENSUSAT	S000407	CANNTG	MYC binding site in the promoters of many dehydration-responsive genes in *Arabidopsis*.
NODCON1GM	S000461	AAAGAT	Promoter motif identified from soybean that is required for nodule-specific expression.
NODCON2GM	S000462	CTCTT	Promoter motif identified from soybean that is required for nodule-specific expression.
OSE1ROOTNODULE	S000467	AAAGAT	Promoter element that is required for gene expression in infected root nodule cells of *Vicia faba*.

OSE2ROOTNODULE	S000468	CTCTT	Element identified in Vicia faba as being required for gene expression infected root nodule cells.
POLLEN1LELAT52	S000245	AGAAA	One of two co-dependant elements responsible for pollen specific expression of the tomato *lat52* gene.
PYRIMIDINEBOXOSRAMY1A	S000259	CCTTTT	Pyrimidine box identified in the rice *RAmy1A* gene promoter as being involved in sugar repression. Also identified in the barley *Amy2/32b* gene promoter as being gibberellin-responsive.
RAV1AAT	S000314	CAACA	Binding site of RAV1 in *Arabidopsis RAV1-A* and *RAV1-B* genes, which are expressed specifically in rosette leaves and roots.
REALPHALGLHCB21	S000362	AACCAA	Motif found in *Lemna gibba Lhcb21* gene promoter, which is required for phytochrome regulation.
ROOTMOTIFTAPOX1	S000098	ATATT	Element found in the *rolD* gene promoter in *Nicotiana tabacum* that is responsible for high root gene expression.
S1FBOXSORPS1L21	S000223	ATGGTA	Identified in the spinach *RPS1* and *RPL21* gene promoters and functions as a negative regulator of transcription.
SEF1MOTIF	S000006	ATATTTAWW	Promoter binding domain for SEF1 (soybean embryo factor 1) from soybean.
SEF4MOTIFGM7S	S000103	RTTTTTR	Promoter binding domain for SEF4 (soybean embryo factor 4) from soybean.
TAAAGSTKST1	S000387	TAAAG	Motif found in *KST1* gene promoter from *Solanum tuberosum*. Target site for StDof1 protein controlling guard cell-specific gene expression.
WBOXATNPR1	S000390	TTGAC	Motif found in the promoter of the *Arabidopsis NPR1* gene, which is recognised by salicylic acid-induced WRKY DNA binding proteins.
WBOXHVISO1	S000442	TGACT	Sugar-responsive promoter motif identified in the barley *iso1* gene promoter.
WBOXNTERF3	S000457	TGACY	Promoter element found in the *ERF3* gene promoter in tobacco, which may be involved in ERF3 gene activation in response to wounding.
WRKY71OS	S000447	TGAC	Binding site of rice WRKY71, a transcriptional repressor of the gibberellin-signalling pathway.

Table 5.3: Suggested roles of the CAREs common to the proline rich protein and chalcone synthase promoters. CAREs that are common to all three of the promoters are indicated in red and those that are common to only the PRP and CHS promoters are indicated in black. The PLACE site identity numbers, sequence motifs and suggested functions of the CAREs are shown.

promoters of the PRP and CHS genes. Analysis was limited to the first 500 bp of promoter upstream of the putative start site of transcription because motifs that direct tissue-specific expression are usually located within the first several hundred base pairs upstream of the start site of transcription (Tyagi, 2001) and because inclusion of the full promoter sequences would lead to the identification of too many shared sequences by chance alone.

This comparison identified 17 common regions of at least seven nucleotides in length (table 5.4). Four of these motifs were also identified in the GBSS promoter (table 5.4 and figure 5.12) but they were all located more than 600 nucleotides upstream of the presumed start site of transcription, suggesting that they are probably not involved in high level, tissue-specific transcription. These motifs are also rich in adenine and thymine and, given the adenine- and thymine-rich nature of the promoters, their presence is likely due to chance.

Five of the novel motifs were present within the first 300 bp upstream of the putative start site of transcription of both promoters, which could suggest their involvement in the regulation of high-level gene transcription (table 5.4 and figures 5.9 and 5.10). However, none of these novel motifs were identified within the GBSS promoter.

The dot plot analysis has significant shortcomings when applied to only three promoters. The small number of promoters available means that the common motifs may occur in three promoters by chance alone. Another consequence of the low number of available promoters is that the identification of short motifs was not possible, as too many false positives would be expected. In addition, CAREs often possess ambiguous consensus sequences, but the analysis performed here can only identify motifs with perfect homology due to the increased number of false positives that would be obtained by searching for ambiguous CAREs.

5.4: Summary and conclusions

In previous chapters, the identification and isolation of three promoters that drive transcription of mRNA transcripts specifically or preferentially in the boll wall were described. These promoters were cloned into a reporter vector, which was used to transiently express the reporter gene in various tissues. The PRP and CHS promoters drove expression in the boll wall between 5 and 25 DPA and 3 and 25 DPA, respectively. Both promoters drove expression in fibre tissue and the CHS promoter also drove expression in petals. Expression of the reporter was not seen in leaf, bract or calyx tissues for either promoter or in petal tissue for the PRP promoter. The expression patterns of the reporter gene driven by the isolated

Motif ID	Position in PRP	Position in CHS	Position in GBSS	Sequence	Length
1	-432	-481	-799	TAAAAAT	7
2	-423	-420		ATATTTAA	8
3	-443	-417		TTTAAGCT	8
4	-274	-366		CAACATT	7
5	-412	-319		CATAGTT	7
6	-423	-307	-832	ATATTTA	7
7	-265	-305		ATTTAGAA	8
8	-215	-300		GAAGTTA	7
9	-345	-264		TTTTTAAG	8
10	-201	-256		TTTGTTTC	8
11	-23	-251		TTCACCATCT	10
12	-434	-244	-602 & -737 & -818	TCTAAAA	7
13	-423	-236		ATATTTAAT	9
14	-405	-212	-649	TTGAAAT	7
15	-415	-171		TTACATAGT	9
16	-288	-103		ATTAGAT	7
17	-42	-66		CTCCCAA	7

Table 5.4: Identification of novel promoter motifs.
The novel motifs that were identified in the PRP and CHS promoters by dot plot analysis are shown above. Motifs present within the first 300 bp upstream of the putative start site of transcription in the PRP and CHS promoters are indicated in orange. The motifs that were also identified in the GBSS promoter are indicated. The position of CAREs within each promoter, relative to the presumed start site of transcription, as well as the sequence and length of each motif are shown.

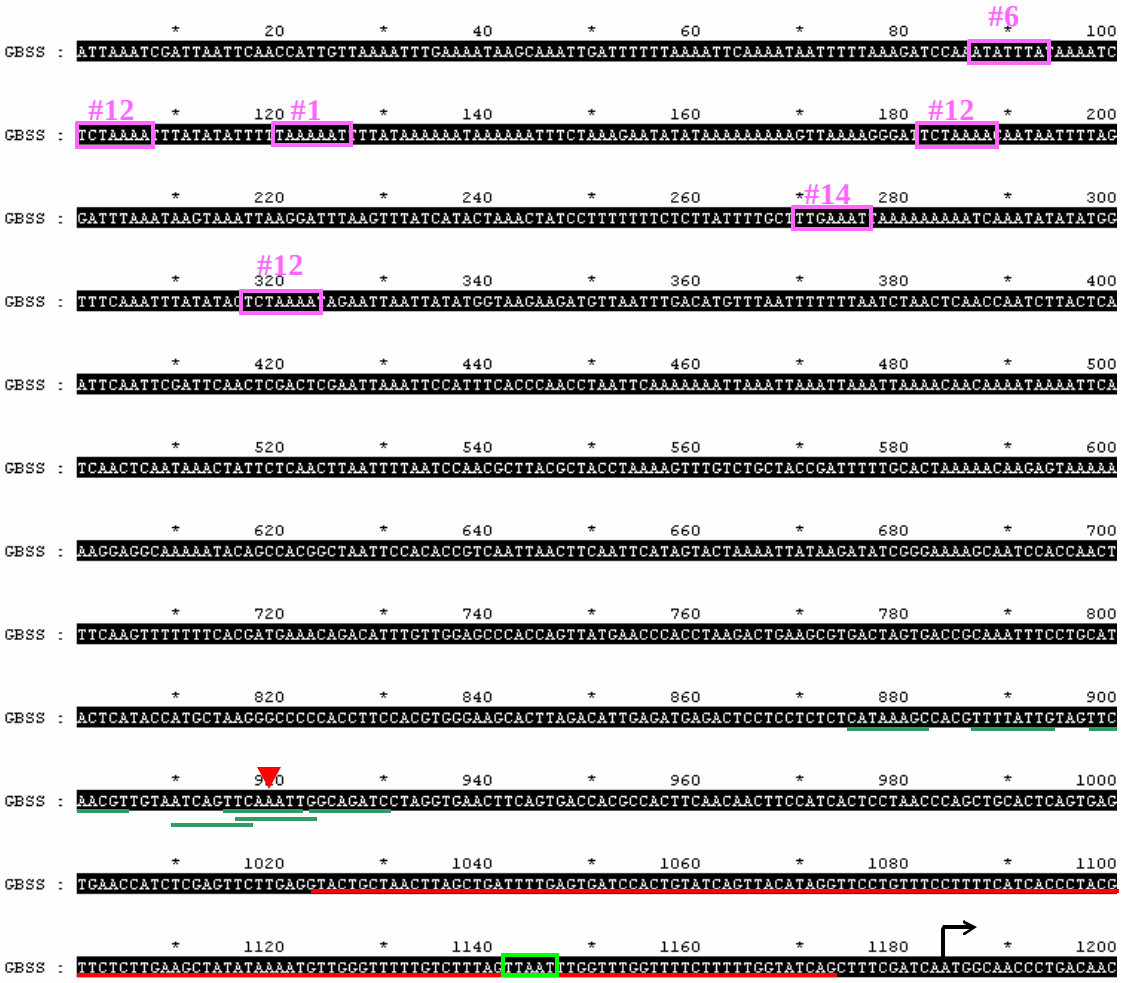


Figure 5.12: Granule-bound starch synthase promoter region.
The presumptive GBSS promoter region and the 5' end of the GBSS mRNA transcript are shown in the alignment above. The putative start sites of translation and transcription are indicated with a black arrow and a red triangle, respectively. The intron that was omitted from the GBSS promoter is underlined in red and a green box shows a potential intron branch point. Potential pyrimidine-rich initiator (Inr) elements, near the start site of transcription, are underlined in green, and pink boxes indicate novel potential CAREs that were present in the PRP, CHS and GBSS promoters.

promoter fragments were similar to the *in planta* transcription patterns, indicating that the promoter fragments isolated contain the CAREs required for *in planta* expression. However, transient transformations with the third promoter, GBSS, failed to yield any positive transformation, probably due to the amplification of an incomplete promoter.

Analysis of promoter sequences for known CAREs revealed the presence of many elements but there was no evidence to suggest the involvement of any particular CARE in boll wall gene expression. The database contains only 453 motifs with few cotton sequences, so the inconclusive results from the analysis are not surprising. It is likely that appropriate CAREs are not contained within this database and therefore a search for novel CAREs that are common to the promoters was performed. 17 putative CAREs, of at least seven nucleotides in length, were identified within the 500 bp upstream of the start sites of transcription of the PRP and CHS promoters, with four of these motifs also identified in the GBSS promoter.

The computational analysis described here is severely constrained by the small number of promoter sequences. It is possible that the motifs identified are involved in the transcription of high levels of the mRNAs in the boll wall, but confirmation is required. The addition of more functionally verified promoter sequences (e.g. GBSS, ANR, DFR, SuSy) would greatly improve the power of the analysis.

Chapter 6: Discussion and conclusions

6.1: Introduction

In recent years, cotton varieties engineered with transgenes derived from the soil bacterium *Bacillus thuringiensis* have been commercialised. The *Bt*-transgenes are under the control of modified *35S* promoters and confer resistance to key cotton pests, with the transgenic cultivars showing improved resistance to insect attack. In addition to improved pest control, *Bt*-cotton has allowed for a reduction in pesticide usage, resulting in financial and environmental benefits. However, while these transgenic varieties have been successful and widely adopted, the constitutive transgene expression may not be necessary since the major Australian cotton pests preferentially attack the flower and boll structures of the plant causing minor, if any, damage to other plant structures. Constitutive high level transgene expression may have several detrimental effects including placing strong selective pressure on pest populations to develop resistance, non-target effects of the transgene, a yield penalty to the plant and the presence of transgenic protein in secondary commercial products. These problems could be avoided by using a tissue-specific promoter to drive transgene expression in only the boll wall.

6.2: Summary of results and implications

The aim of this study was to identify and characterise cotton tissue-specific promoters that drive expression in only the boll wall. In order to identify promoters that drive boll wall specific expression, mRNA transcripts that are highly abundant and specific to the boll wall were identified and the corresponding promoters isolated. Transcripts that are more abundant in 5 DPA boll wall than leaf tissue were identified using a differential screening approach. Six major classes of cDNA were identified from this screen and correspond to a PRP, CHS, SuSy, DFR, ANR, GBSS, δ-TIP and actin. According to publicly available databases, the PRP, CHS, DFR and GBSS sequences identified correspond to cotton genes that have not been previously characterised.

The transcription patterns of the putative boll wall-specific mRNAs were investigated in a variety of tissue types and developmental timepoints of the boll wall using Northern analysis. Transcripts corresponding to δ-TIP and actin were present at significant levels in tissues other than the boll wall and the corresponding promoters were therefore deemed unsuitable for the aims of this project. The PRP, CHS, SuSy, DFR, ANR and GBSS transcripts were detected at relatively high levels during the early stages of boll wall development (0 to 15 DPA), with decreasing abundance as the boll matures. Minor levels of

several of these transcripts were also detected in tissues other than the boll wall. The transcription patterns of the PRP, CHS and GBSS transcripts indicated that the corresponding promoters are likely to be of most interest to this study, as the transcripts are highly abundant in the boll wall and have the most restricted spatial patterns.

The promoters corresponding to these three transcripts were identified using Genome Walker® PCR and cloned into the promoterless pJK vector, upstream of the *gusA* reporter gene, to facilitate an analysis of their functional specificity. Transient transformation of various cotton tissues, namely boll wall (3, 5, 10, 20 and 30 DPA), leaf, bract, calyx, fibre and petal, were performed with the promoter:reporter constructs.

Biolistic analysis with the *PRP::GUS* construct resulted in GUS-expressing foci in the boll wall at 5, 10, 20 and 25 DPA as well as in fibre. The *CHS::GUS* construct gave expression at 3, 5, 10, 20 and 25 DPA in the boll wall as well fibre and petal tissues. The GUS expression patterns driven by the PRP and CHS promoters correlates well with their transcript abundance as demonstrated by Northern analysis. This similarity indicates that both of the promoter fragments contain the necessary elements for the boll wall-preferential gene transcription that is seen *in planta*. However, transformation with the *GBSS::GUS* construct gave no detectable expression in any of the tissues tested. This lack of expression is likely to be the result of a technical problem, which resulted in the omission of a 5' UTR intronic sequence from the amplified promoter sequence.

While the PRP and CHS promoters did not drive reporter expression specifically in the boll wall, reporter protein was present in high levels in the boll wall, and only minor expression was seen in fibre and flower tissues. Both of these tissues account for a small proportion of the total biomass of the plant and are not tissues that are used in secondary commercial products, meaning that the non-target expression driven by these promoters may not necessarily be a problem. The flower expression of the reporter protein driven by the CHS promoter may actually be a useful feature of the promoter, as the key Australian cotton pests attack the flower as well as the boll. Thus it may be beneficial for future transgenic varieties to express anti-pest molecules in flower tissues in addition to the boll wall. Overall, these observations suggest that the PRP and CHS promoters may be useful in the development of future transgenic varieties.

6.3: Future experiments

While this study identified two promoters that drive high-level boll wall-preferential expression, the functional analysis of a third promoter (that of the GBSS transcript) was not successful. The GBSS transcript was detected in the boll wall at later timepoints than the other transcripts, and therefore its promoter may be of more use for expressing anti-pest molecules than the other two promoters identified here. Thus, the isolation of a more complete GBSS promoter and testing of its functional specificity may be worthwhile to undertake in the future.

Relative to the start site of translation, the isolated PRP, CHS and GBSS promoter fragments are respectively 1582 bp, 708 bp and 1183 bp (including the 151 bp intron) in length. The only major discordance between the Northern analysis and the transient transformation experiments is for CHS in the flower tissue. The CHS transcript is highly abundant in petal and early boll wall timepoints, but while the isolated CHS promoter drove a high level of reporter expression in the early boll wall timepoints, expression in petal tissues was relatively low. It seems likely that the reduced reporter expression in petals is due to the isolation of a relatively short CHS promoter fragment. If this assumption is correct, it would be beneficial to identify more 5' promoter sequence using the Genome Walker® PCR method described in chapter 4.

Ultimately, to determine whether these promoters might be effective in future transgenic cultivars, it would be necessary to generate transgenic cotton plants with the promoters driving anti-pest transgene expression. Using field trials, insect damage could be compared between the new lines with tissue-specific transgene expression and current transgenic cultivars with constitutive endotoxin expression. However, given the time and cost involved in the generation of transgenic cotton and field trials, it may be prudent to further characterise and develop the promoters before generating transgenic plants.

The first step in further development of the promoters is the identification and functional analysis of more complete CHS and GBSS promoters. Following this, a deletion analysis of the promoters using successive 5' deletions, combined with functional testing of the promoter fragments, could be undertaken to identify elements or regions that are responsible for boll wall or flower expression. This type of analysis has been previously successful in identifying short promoter regions critical for tissue-specific expression in cotton (Harmer, 2003, Delaney, 2005).

Promoters used in transgenic cultivars are often extensively modified from their original composition in order to drive the desired expression pattern. In transgenic cotton cultivars, the *35S* promoters used to drive the *cry1Ac* and *cry2Ab* genes are enhanced and tandemly repeated. The *cry2Ab* promoter also contains leader sequences from the petunia *heat shock 70* gene[10]. Similar modifications, such as the addition or removal of CAREs or regions of sequence, or the alteration of the nucleotide composition, of the promoters identified in this study may optimise their performance. Furthermore, novel synthetic promoters could be designed based upon the identification of CAREs in these promoters that drive boll wall and/or flower specific expression.

After further characterisation and optimisation of these tissue-specific promoters, transgenic plants with one of the promoters driving expression of an anti-pest molecule could be generated. While any gene with anti-pest properties could be used to test the performance of these promoters, it would be beneficial to use the *cry* genes from *B. thuringiensis* because they are well characterised, have regulatory approval and allow for a comparison against the established transgenic cotton cultivars. Numerous transgenic lines would be required to enable selection of a suitable transformation event with high-level transgene expression, as expression can vary depending upon the insertion site of the transgene.

Transgenic lines containing the promoter driving expression of an anti-pest gene would then be compared in field trials to current transgenic varieties with constitutive transgene expression. Such testing would reveal whether anti-pest transgene expression in only the flower and/or boll wall is indeed sufficient to provide adequate protection from insect attack. For example, the levels of transgene expression throughout the plant and rates of insect survival should be measured. It would also be important to compare the efficacy of new transgenic lines over time with the current Bollgard II® varieties, which show declining efficacy with plant age (Adamczyk *et al.*, 2001b). Rates of insect survival on pollen should also be measured, as it has been noted that *H. armigera* are able to survive on Bollgard II® by feeding upon the pollen of the plant, which has relatively low endotoxin expression (Greenplate, 1999). Declining efficacy with plant age and relatively low transgene expression in the pollen are undesirable, as they lead to insects receiving a sub-lethal dose of endotoxin, which may promote the development of resistance within the insect population. Therefore, new transgenic varieties may represent an improvement upon these aspects of current varieties.

[10] The Office of the Gene Technology Regulator (http://www.ogtr.gov.au)

While the promoters isolated in this study did not drive reporter expression in boll wall timepoints after 25 DPA, development of a modified or synthetic promoter may produce one that drives expression at later timepoints. However, transgene transcription persisting until later in development may not actually be necessary, as stability of the transgenic protein may ensure high levels of endotoxin persist until later in boll wall development. In addition, as the boll matures it becomes less attractive as a food source to pests (Pyke and Brown, 1996).

An important consideration for future transgenic varieties with tissue-specific anti-pest transgene expression is the potential for the development of resistance in the pest population. As is the case with the current transgenic cultivars, new cultivars with tissue-specific endotoxin expression should be be fatal to a very high proportion of homozygous and heterozygous susceptible insects. Measurements of any non-target endotoxin expression would also need to be monitored in new varieties, and the implications of this expression on pest population resistance considered. Conditions under which insects are exposed to a sub-lethal dose of endotoxin may select for resistance alleles within the population (Frutos *et al.*, 1999). This may be overcome with promoter modification designed to remove non-target expression.

6.4: Conclusions

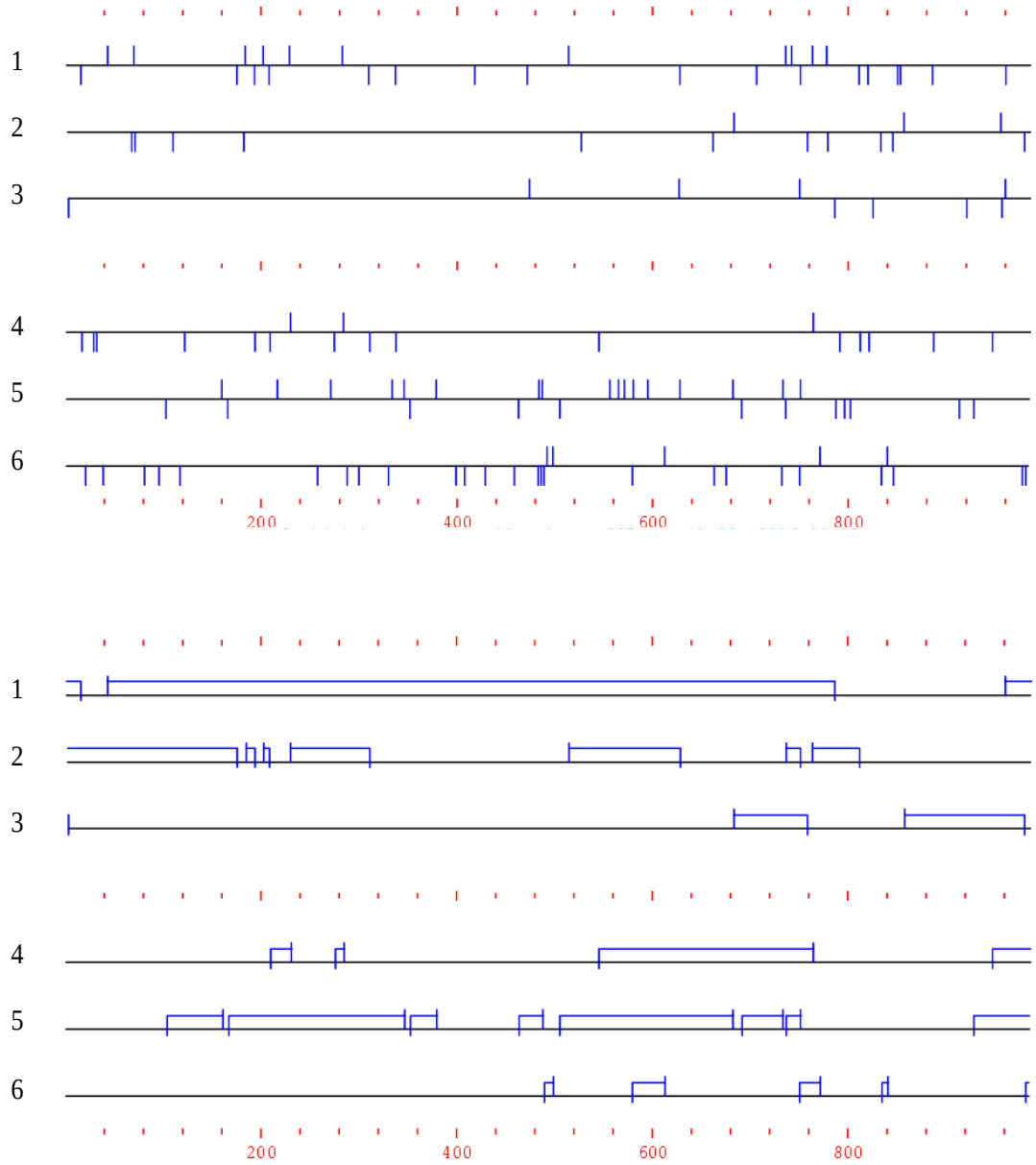
The aims of this project have been largely achieved, with the identification of two promoters that drove high-level reporter expression preferentially in the boll wall. Transient transformations with promoter:reporter constructs revealed that the PRP promoter drove expression in 5 to 25 DPA boll wall and fibre tissues while the CHS promoter drove expression in 3 to 25 DPA boll wall, fibre and petal tissues.

The *Bt*-transgenic cotton cultivars that were first commercially released in 1996 were based on research that was conducted more than 15 years ago. Since then, more sophisticated methods for transgene expression using inducible, tissue-specific, time-specific or novel synthetic promoters have become available (reviewed by Schuler *et al.*, 1998; Lessard *et al.*, 2002; Gurr and Rushton, 2005). While current *Bt*-transgenic varieties have been successful in Australia, resulting in increased protection from insect attack and improved environmental impacts, future transgenic cultivars can be further improved. In the research presented here, promoters that could be utilised in the next generation of transgenic cotton cultivars have been identified and characterised.

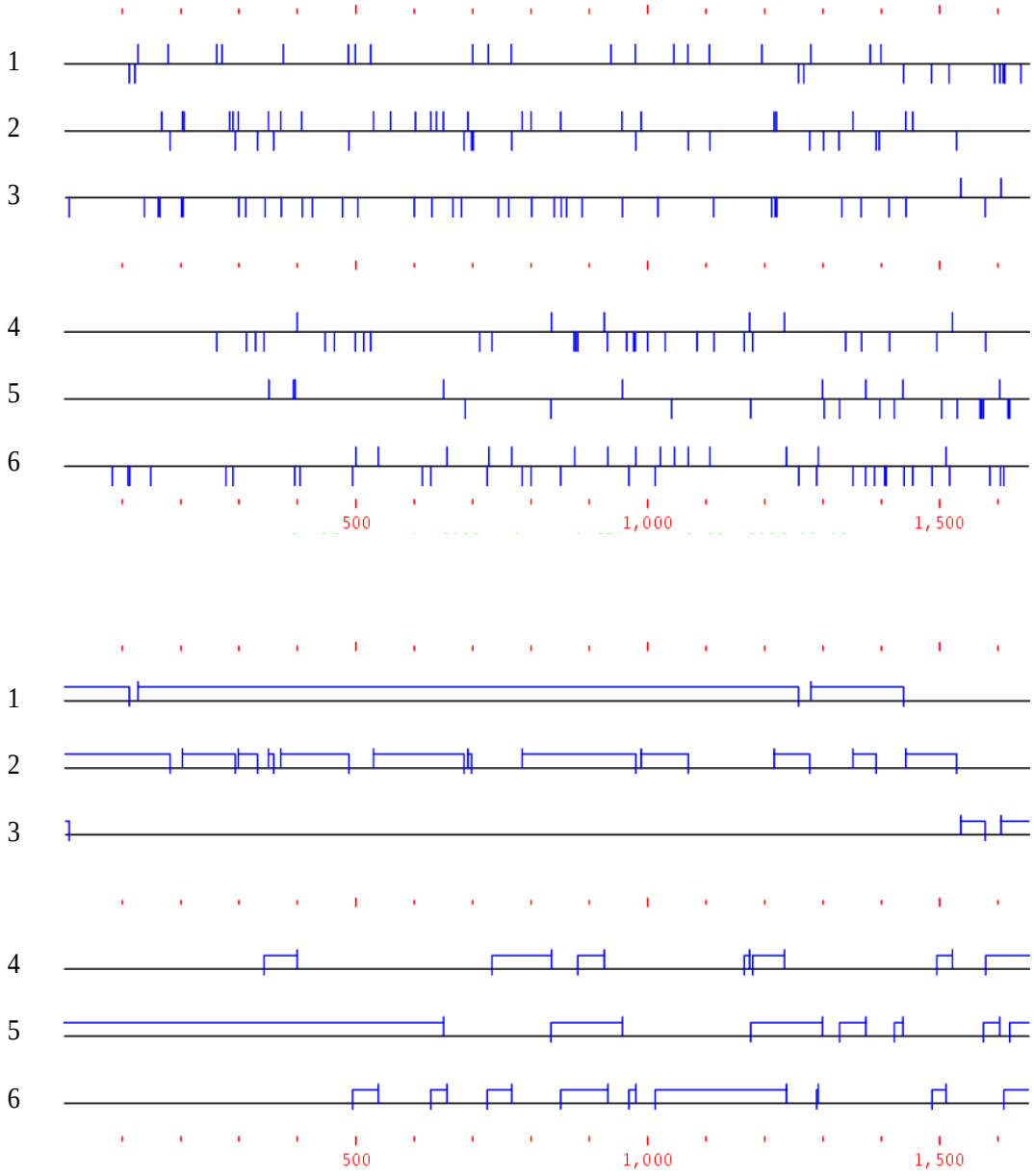
Appendix 1: Potential open reading frames and translational start and stop codons identified within the eight analysed cDNA sequences

Each of the appendices (A-H) contains two panels, with the top panel representing potential translational start and stop codons identified within the cDNA and the bottom panel representing potential open reading frames. The six potential reading frames are indicated 1 to 6 on the left of the figure, with the 5’ direction reading frames indicated by numbers 1 to 3 and the 3’ reading frames indicated by numbers 4 to 6. A horizontal black line represents the cDNA sequence with 100 bp intervals marked in red below the figure. The top panel in each of appendices shows potential start and stop codons identified respectively as vertical lines above or below the cDNA. The bottom panel shows potential open reading frames within each of the six reading frames of each cDNA.

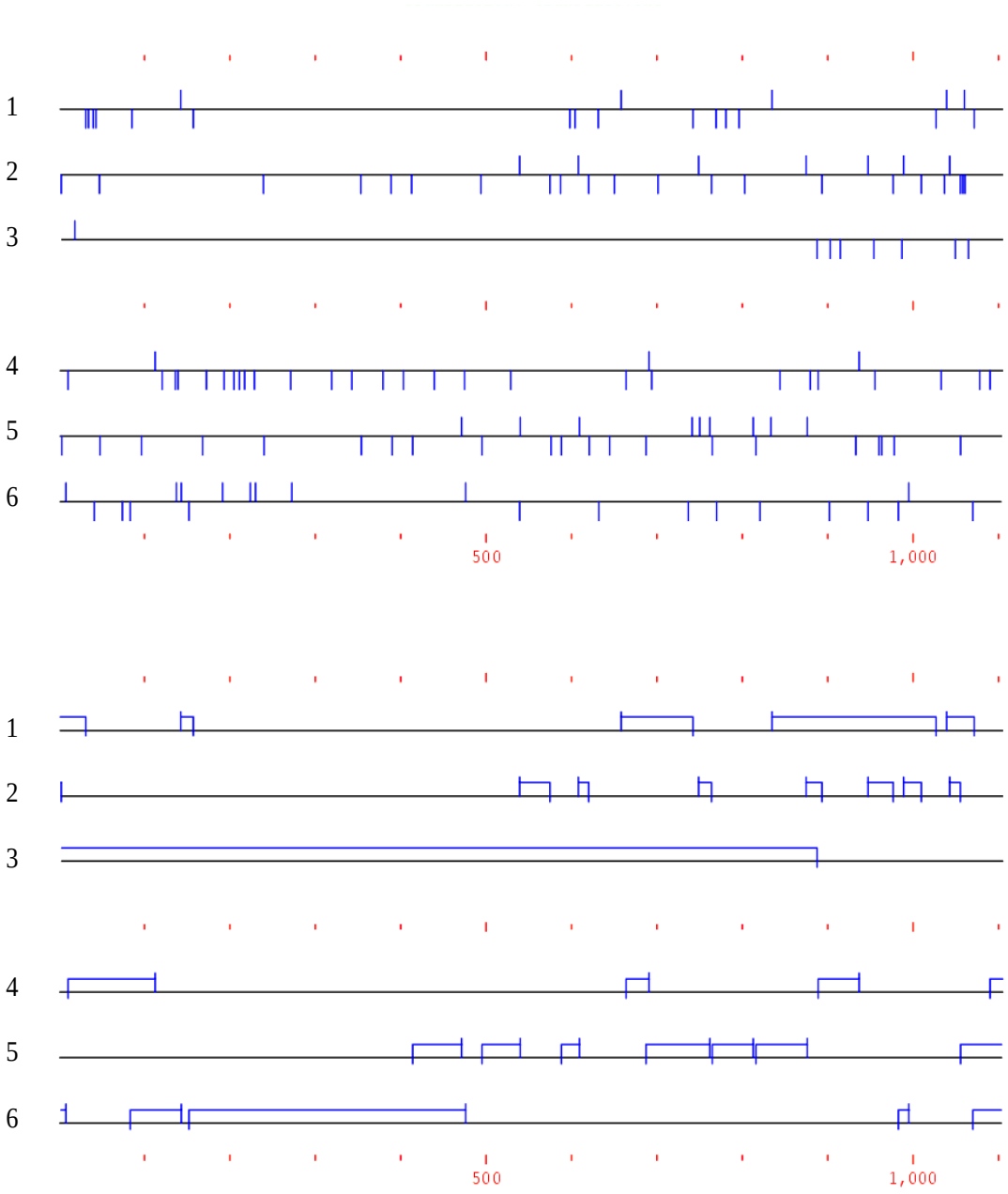
Appendix 1A: <u>Potential open reading frames and translational start and stop codons identified within the δ-tonoplast intrinsic protein cDNA</u>



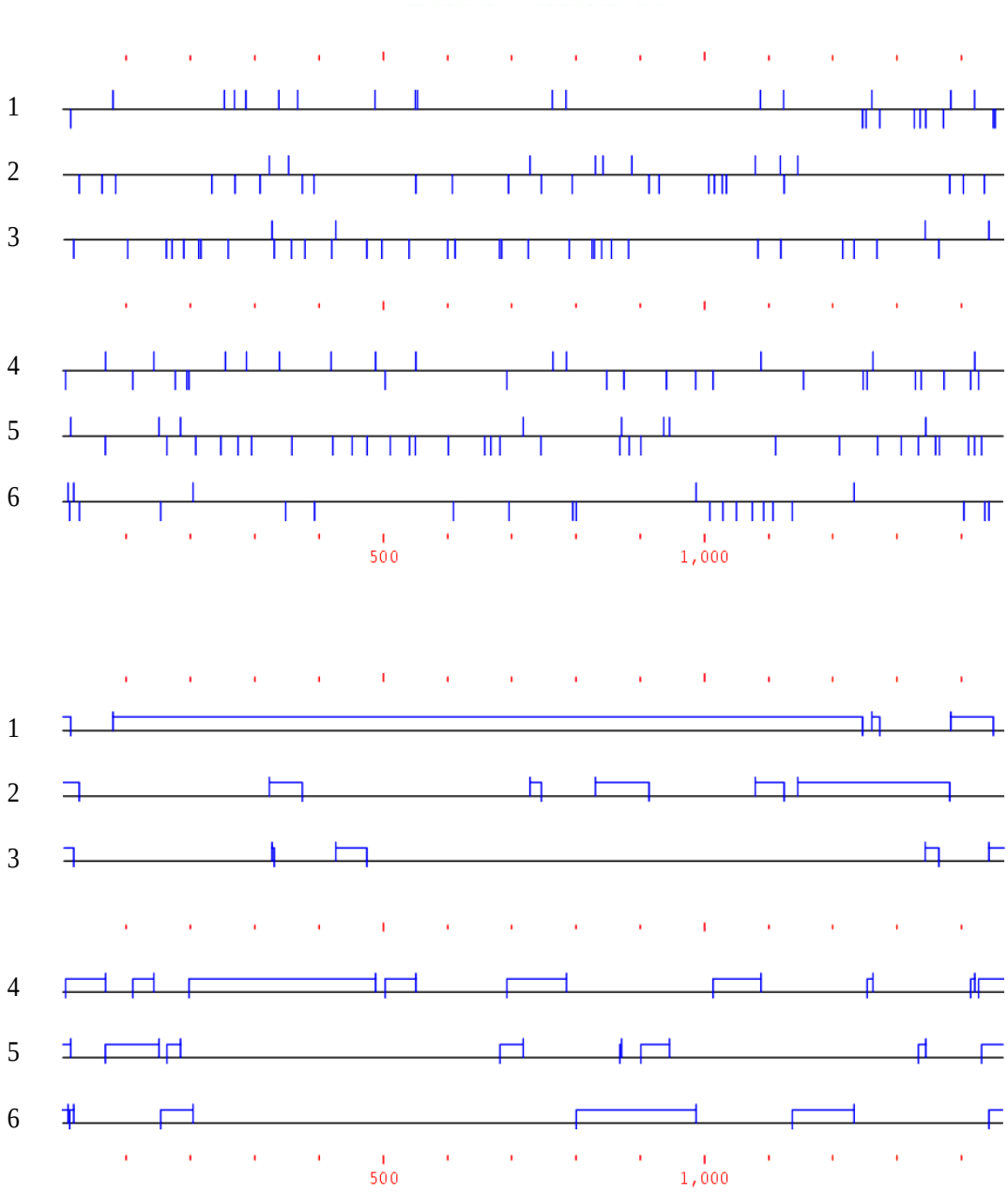
Appendix 1B: <u>Potential open reading frames and translational start and stop codons identified within the actin cDNA</u>



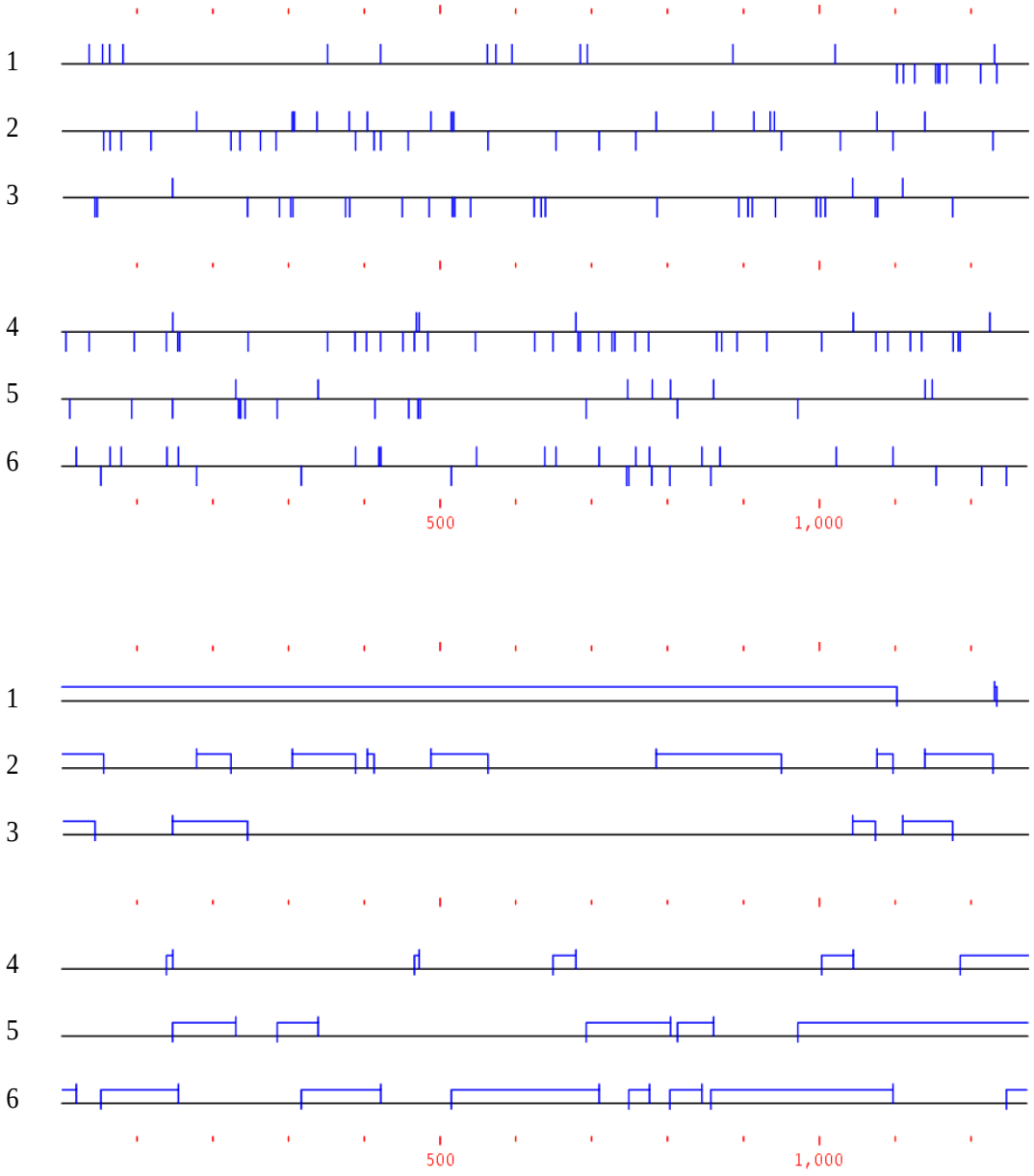
Appendix 1C: <u>Potential open reading frames and translational start and stop codons identified within the proline rich protein cDNA</u>



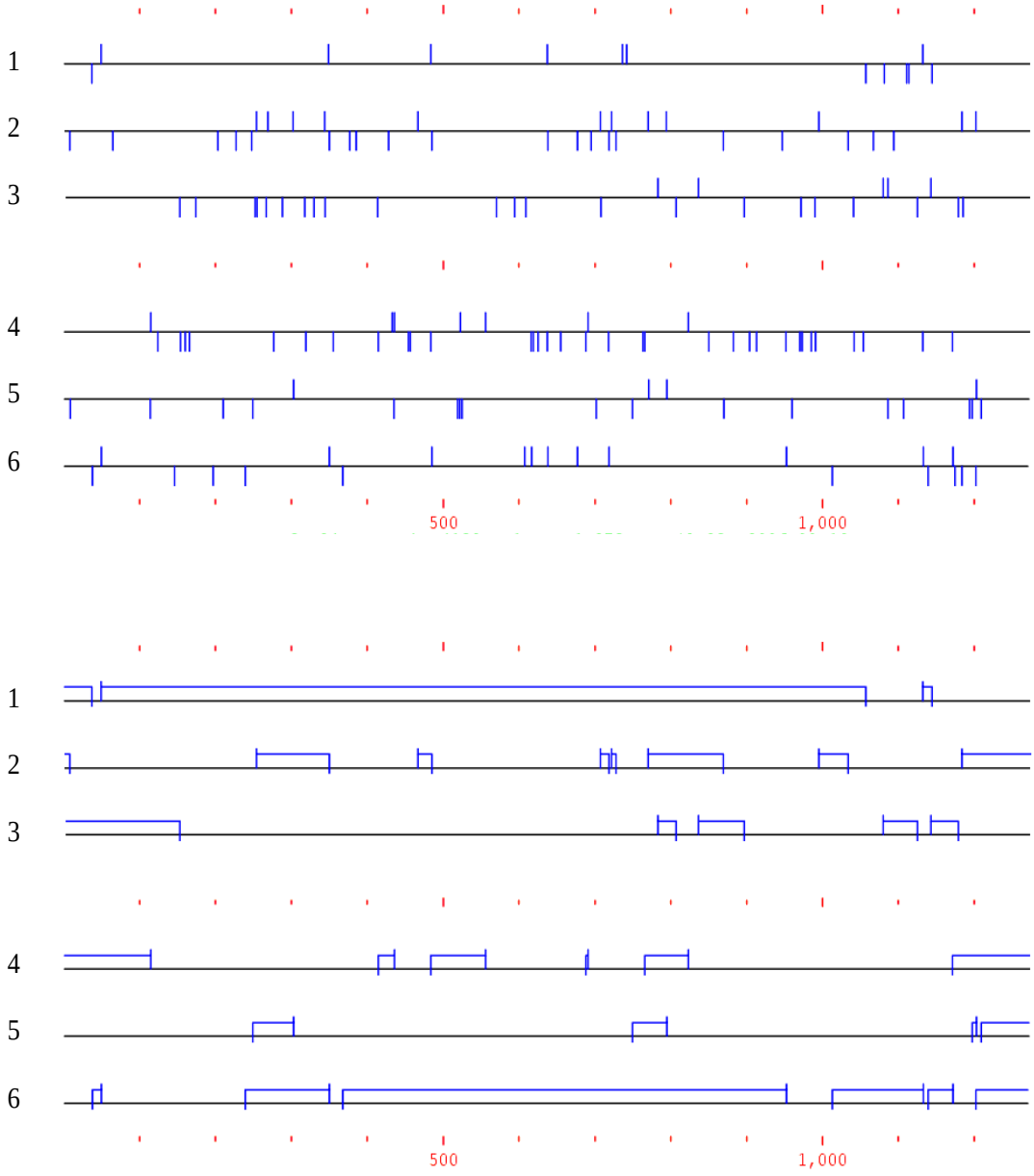
Appendix 1D: <u>Potential open reading frames and translational start and stop codons identified within the chalcone synthase cDNA</u>



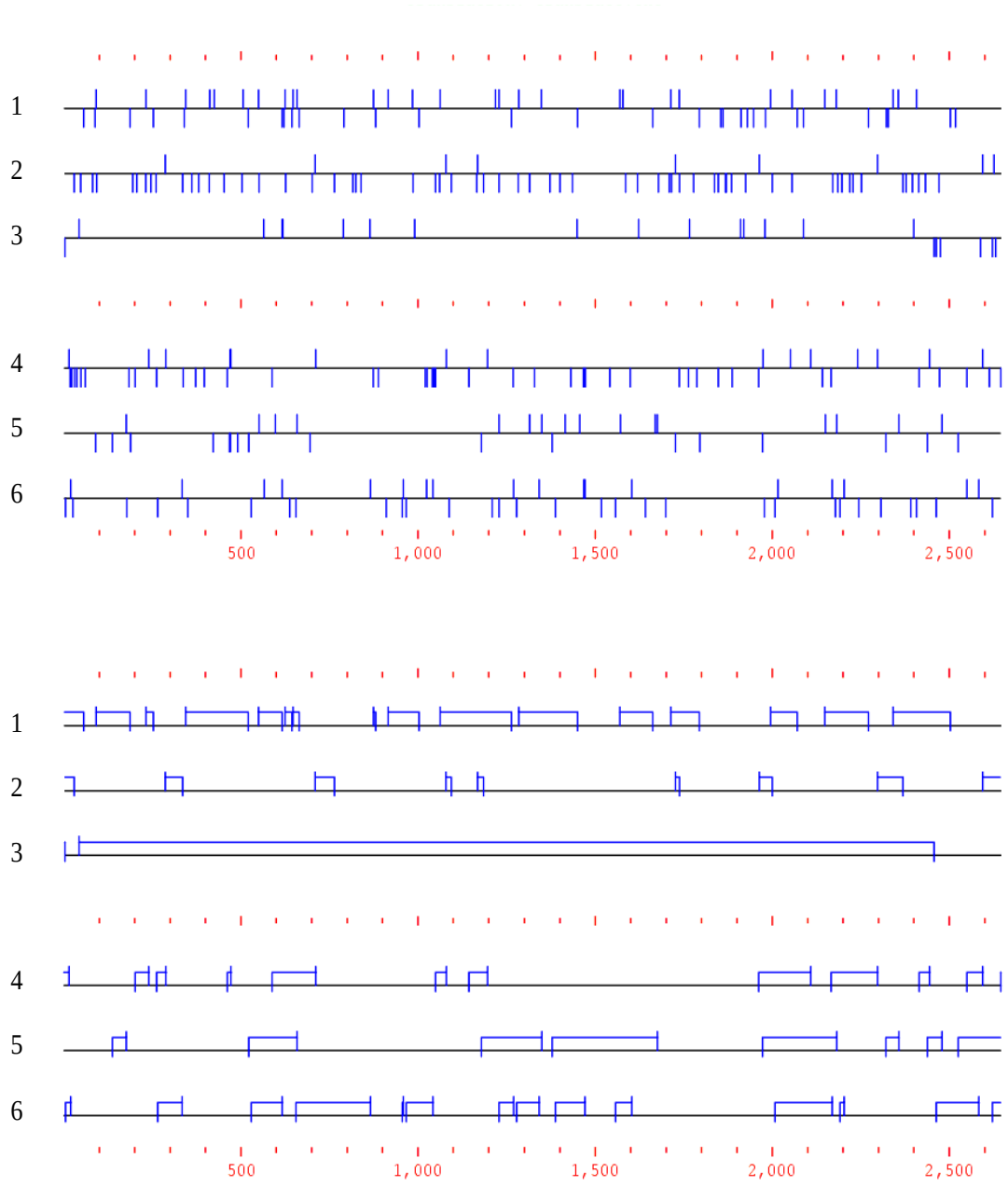
Appendix 1E: <u>Potential open reading frames and translational start and stop codons identified within the dihydroflavonol 4-reductase cDNA</u>



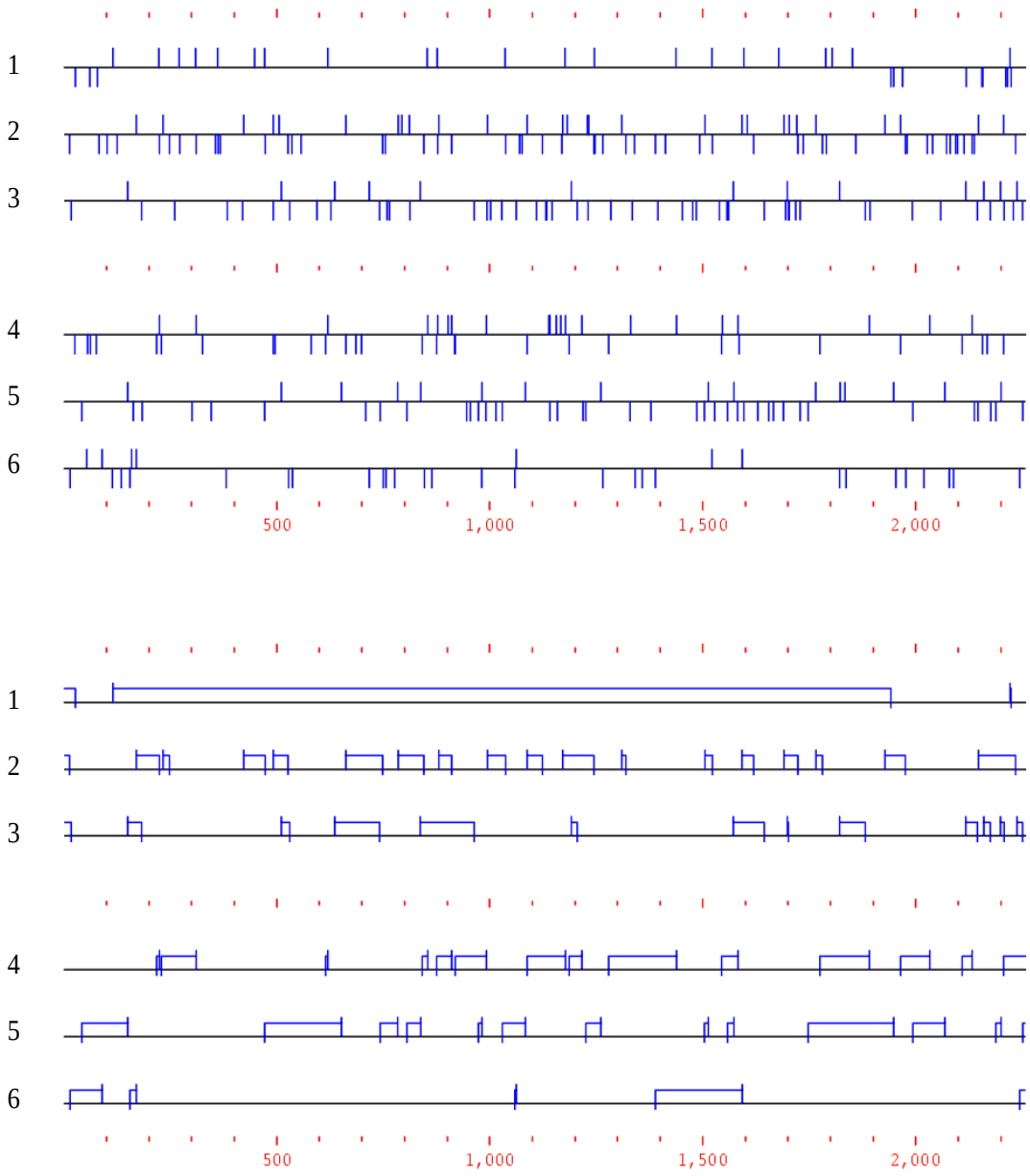
Appendix 1F: Potential open reading frames and translational start and stop codons identified within the anthocyanidin reductase cDNA



Appendix 1G: <u>Potential open reading frames and translational start and stop codons identified within the sucrose synthase cDNA</u>



Appendix 1H: <u>Potential open reading frames and translational start and stop codons identified within the granule-bound starch synthase cDNA</u>



Appendix 2: Plant *cis*-acting regulatory DNA elements identified within the three analysed promoters

Each of the appendices (A-C) shows the PLACE motifs identified within the PRP, CHS and GBSS promoters. The PLACE motif name, location, sequence and PLACE database identification numbers for each motif are shown. The CAREs located within the putative intron in the GBSS promoter are indicated in red in appendix 2C.

Appendix 2A: PLACE motifs identified within the proline rich protein promoter

Site Name	Location	Strand	Sequence	SITE ID
DOFCOREZM	9	(-)	AAAG	S000265
POLLEN1LELAT52	10	(-)	AGAAA	S000245
ARR1AT	15	(-)	NGATT	S000454
DOFCOREZM	27	(-)	AAAG	S000265
POLLEN1LELAT52	28	(-)	AGAAA	S000245
CARGCW8GAT	31	(+)	CWWWWWWWWG	S000431
CARGCW8GAT	31	(-)	CWWWWWWWWG	S000431
CACTFTPPCA1	47	(-)	YACT	S000449
GTGANTG10	54	(+)	GTGA	S000378
ARR1AT	55	(+)	NGATT	S000454
WBOXATNPR1	77	(+)	TTGAC	S000390
WRKY71OS	78	(+)	TGAC	S000447
BIHD1OS	78	(-)	TGTCA	S000498
ARR1AT	86	(+)	NGATT	S000454
CCAATBOX1	95	(+)	CCAAT	S000030
CAATBOX1	96	(+)	CAAT	S000028
ARR1AT	97	(-)	NGATT	S000454
GT1CONSENSUS	103	(+)	GRWAAW	S000198
POLASIG2	106	(+)	AATTAAA	S000081
PYRIMIDINEBOXOSRAMY1A	110	(-)	CCTTTT	S000259
DOFCOREZM	111	(+)	AAAG	S000265
SURECOREATSULTR11	115	(+)	GAGAC	S000499
GTGANTG10	130	(-)	GTGA	S000378
DOFCOREZM	134	(-)	AAAG	S000265
DOFCOREZM	142	(-)	AAAG	S000265
TAAAGSTKST1	142	(-)	TAAAG	S000387
CACTFTPPCA1	152	(+)	YACT	S000449
DOFCOREZM	154	(-)	AAAG	S000265
POLLEN1LELAT52	156	(-)	AGAAA	S000245
BOXIINTPATPB	157	(-)	ATAGAA	S000296
CACTFTPPCA1	165	(-)	YACT	S000449
RAV1AAT	170	(-)	CAACA	S000314
GT1CONSENSUS	177	(-)	GRWAAW	S000198
SREATMSD	178	(+)	TTATCC	S000470
IBOXCORE	178	(-)	GATAA	S000199
GATABOX	179	(-)	GATA	S000039
MYBST1	179	(-)	GGATA	S000180
CAATBOX1	195	(+)	CAAT	S000028
ARR1AT	196	(-)	NGATT	S000454
CATATGGMSAUR	199	(+)	CATATG	S000370
EBOXBNNAPA	199	(+)	CANNTG	S000144
MYCCONSENSUSAT	199	(+)	CANNTG	S000407
CATATGGMSAUR	199	(-)	CATATG	S000370
EBOXBNNAPA	199	(-)	CANNTG	S000144
MYCCONSENSUSAT	199	(-)	CANNTG	S000407
RYREPEATGMGY2	202	(-)	CATGCAT	S000105
RYREPEATLEGUMINBOX	202	(-)	CATGCAY	S000100
RYREPEATBNNAPA	203	(-)	CATGCA	S000264
BIHD1OS	211	(+)	TGTCA	S000498
WRKY71OS	212	(-)	TGAC	S000447
NODCON2GM	224	(+)	CTCTT	S000462
OSE2ROOTNODULE	224	(+)	CTCTT	S000468
DOFCOREZM	226	(-)	AAAG	S000265
TAAAGSTKST1	226	(-)	TAAAG	S000387
WBOXHVISO1	233	(-)	TGACT	S000442
WBOXNTERF3	233	(-)	TGACY	S000457
WRKY71OS	234	(-)	TGAC	S000447
CACTFTPPCA1	238	(+)	YACT	S000449
PREATPRODH	239	(+)	ACTCAT	S000450
INRNTPSADB	240	(+)	YTCANTYY	S000395
GATABOX	253	(+)	GATA	S000039
IBOXCORE	253	(+)	GATAA	S000199
MYBCORE	255	(-)	CNGTTR	S000176
MYBCORE	266	(+)	CNGTTR	S000176
GT1CORE	267	(+)	GGTTAA	S000125
CAATBOX1	273	(-)	CAAT	S000028
ARR1AT	275	(+)	NGATT	S000454

CAATBOX1	277	(-)	CAAT	S000028
CCAATBOX1	277	(-)	CCAAT	S000030
RYREPEATLEGUMINBOX	281	(-)	CATGCAY	S000100
RYREPEATBNNAPA	282	(-)	CATGCA	S000264
EBOXBNNAPA	284	(+)	CANNTG	S000144
MYCATERD1	284	(+)	CATGTG	S000413
MYCCONSENSUSAT	284	(+)	CANNTG	S000407
EBOXBNNAPA	284	(-)	CANNTG	S000144
MYCATRD22	284	(-)	CACATG	S000174
MYCCONSENSUSAT	284	(-)	CANNTG	S000407
GTGANTG10	287	(+)	GTGA	S000378
GATABOX	289	(+)	GATA	S000039
GT1MOTIFPSRBCS	291	(+)	KWGTGRWAAWRW	S000051
CACTFTPPCA1	292	(-)	YACT	S000449
GTGANTG10	293	(+)	GTGA	S000378
GT1CONSENSUS	295	(+)	GRWAAW	S000198
GT1GMSCAM4	295	(+)	GAAAAA	S000453
BOXIINTPATPB	302	(+)	ATAGAA	S000296
POLLEN1LELAT52	304	(+)	AGAAA	S000245
TATABOX5	306	(-)	TTATTT	S000203
POLASIG3	307	(+)	AATAAT	S000088
ARR1AT	314	(+)	NGATT	S000454
CACTFTPPCA1	332	(+)	YACT	S000449
ROOTMOTIFTAPOX1	342	(+)	ATATT	S000098
TAAAGSTKST1	351	(+)	TAAAG	S000387
DOFCOREZM	352	(+)	AAAG	S000265
NODCON1GM	352	(+)	AAAGAT	S000461
OSE1ROOTNODULE	352	(+)	AAAGAT	S000467
ARR1AT	357	(+)	NGATT	S000454
REALPHALGLHCB21	364	(+)	AACCAA	S000362
CCAATBOX1	366	(+)	CCAAT	S000030
CAATBOX1	367	(+)	CAAT	S000028
ARR1AT	368	(-)	NGATT	S000454
CAATBOX1	371	(+)	CAAT	S000028
GATABOX	377	(+)	GATA	S000039
CACTFTPPCA1	382	(-)	YACT	S000449
GTGANTG10	390	(+)	GTGA	S000378
GATABOX	392	(+)	GATA	S000039
GT1CONSENSUS	392	(+)	GRWAAW	S000198
IBOXCORE	392	(+)	GATAA	S000199
TATABOX5	396	(-)	TTATTT	S000203
AMYBOX1	399	(+)	TAACARA	S000020
GAREAT	399	(+)	TAACAAR	S000439
MYBGAHV	399	(+)	TAACAAA	S000181
AACACOREOSGLUB1	400	(+)	AACAAAC	S000353
MYB1AT	403	(+)	WAACCA	S000408
REALPHALGLHCB21	404	(+)	AACCAA	S000362
ARR1AT	410	(-)	NGATT	S000454
POLLEN1LELAT52	435	(+)	AGAAA	S000245
GT1CONSENSUS	436	(+)	GRWAAW	S000198
POLASIG3	440	(-)	AATAAT	S000088
TATABOX5	441	(+)	TTATTT	S000203
POLLEN1LELAT52	453	(-)	AGAAA	S000245
GT1CONSENSUS	459	(+)	GRWAAW	S000198
DPBFCOREDCDC3	465	(+)	ACACNNG	S000292
CACTFTPPCA1	472	(+)	YACT	S000449
CAATBOX1	486	(-)	CAAT	S000028
GATABOX	489	(+)	GATA	S000039
GT1CONSENSUS	489	(+)	GRWAAW	S000198
IBOXCORE	489	(+)	GATAA	S000199
CAATBOX1	493	(-)	CAAT	S000028
RAV1AAT	523	(+)	CAACA	S000314
GATABOX	537	(-)	GATA	S000039
DOFCOREZM	542	(-)	AAAG	S000265
TAAAGSTKST1	542	(-)	TAAAG	S000387
CACTFTPPCA1	545	(+)	YACT	S000449
DOFCOREZM	569	(-)	AAAG	S000265
ELRECOREPCRP1	576	(-)	TTGACC	S000142
WBOXNTERF3	576	(-)	TGACY	S000457
WBOXATNPR1	577	(-)	TTGAC	S000390
WRKY71OS	577	(-)	TGAC	S000447
CAATBOX1	579	(+)	CAAT	S000028
ROOTMOTIFTAPOX1	580	(-)	ATATT	S000098

ROOTMOTIFTAPOX1	581	(+)	ATATT	S000098
CACTFTPPCA1	622	(+)	YACT	S000449
CURECORECR	629	(+)	GTAC	S000493
CURECORECR	629	(-)	GTAC	S000493
CACTFTPPCA1	630	(+)	YACT	S000449
POLASIG2	635	(+)	AATTAAA	S000081
POLLEN1LELAT52	644	(+)	AGAAA	S000245
WRKY71OS	657	(+)	TGAC	S000447
BIHD1OS	657	(-)	TGTCA	S000498
GT1CONSENSUS	668	(-)	GRWAAW	S000198
IBOXCORE	669	(-)	GATAA	S000199
GATABOX	670	(-)	GATA	S000039
TBOXATGAPB	682	(-)	ACTTTG	S000383
DOFCOREZM	683	(+)	AAAG	S000265
ROOTMOTIFTAPOX1	692	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	693	(+)	ATATT	S000098
TATABOXOSPAL	694	(+)	TATTTAA	S000400
GATABOX	706	(+)	GATA	S000039
GT1CONSENSUS	706	(+)	GRWAAW	S000198
IBOXCORE	706	(+)	GATAA	S000199
ROOTMOTIFTAPOX1	709	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	712	(+)	ATATT	S000098
ARR1AT	724	(-)	NGATT	S000454
CAATBOX1	727	(+)	CAAT	S000028
ARR1AT	728	(-)	NGATT	S000454
LTRE1HVBLT49	733	(-)	CCGAAA	S000250
ARR1AT	738	(+)	NGATT	S000454
TAAAGSTKST1	742	(+)	TAAAG	S000387
NTBBF1ARROLB	742	(-)	ACTTTA	S000273
DOFCOREZM	743	(+)	AAAG	S000265
UP2ATMSD	750	(-)	AAACCCTA	S000472
POLASIG2	755	(-)	AATTAAA	S000081
ARR1AT	763	(+)	NGATT	S000454
CAATBOX1	773	(-)	CAAT	S000028
GATABOX	787	(+)	GATA	S000039
CACTFTPPCA1	790	(-)	YACT	S000449
ARR1AT	812	(+)	NGATT	S000454
CAATBOX1	817	(+)	CAAT	S000028
CACTFTPPCA1	840	(-)	YACT	S000449
CURECORECR	841	(+)	GTAC	S000493
CURECORECR	841	(-)	GTAC	S000493
ACGTABOX	842	(+)	TACGTA	S000130
ACGTABOX	842	(-)	TACGTA	S000130
ACGTATERD1	843	(+)	ACGT	S000415
ACGTATERD1	843	(-)	ACGT	S000415
CAATBOX1	856	(-)	CAAT	S000028
GT1CONSENSUS	859	(+)	GRWAAW	S000198
EECCRCAH1	868	(+)	GANTTNC	S000494
POLLEN1LELAT52	871	(-)	AGAAA	S000245
DOFCOREZM	874	(-)	AAAG	S000265
IBOX	883	(-)	GATAAG	S000124
IBOXCORE	884	(-)	GATAA	S000199
GATABOX	885	(-)	GATA	S000039
DOFCOREZM	896	(+)	AAAG	S000265
NODCON2GM	897	(-)	CTCTT	S000462
OSE2ROOTNODULE	897	(-)	CTCTT	S000468
SURECOREATSULTR11	899	(+)	GAGAC	S000499
ARFAT	899	(-)	TGTCTC	S000270
SEBFCONSSTPR10A	899	(-)	YTGTCWC	S000391
SEF4MOTIFGM7S	903	(-)	RTTTTTR	S000103
TATABOX5	906	(-)	TTATTT	S000203
EBOXBNNAPA	916	(+)	CANNTG	S000144
MYCCONSENSUSAT	916	(+)	CANNTG	S000407
EBOXBNNAPA	916	(-)	CANNTG	S000144
MYCCONSENSUSAT	916	(-)	CANNTG	S000407
CACTFTPPCA1	935	(+)	YACT	S000449
CACTFTPPCA1	949	(+)	YACT	S000449
TBOXATGAPB	950	(+)	ACTTTG	S000383
DOFCOREZM	951	(-)	AAAG	S000265
ARR1AT	954	(+)	NGATT	S000454
GT1CONSENSUS	964	(-)	GRWAAW	S000198
DOFCOREZM	973	(-)	AAAG	S000265
EBOXBNNAPA	978	(+)	CANNTG	S000144

MYCCONSENSUSAT	978	(+)	CANNTG	S000407
EBOXBNNAPA	978	(-)	CANNTG	S000144
MYCCONSENSUSAT	978	(-)	CANNTG	S000407
DOFCOREZM	984	(-)	AAAG	S000265
TAAAGSTKST1	984	(-)	TAAAG	S000387
POLASIG2	985	(-)	AATTAAA	S000081
GTGANTG10	991	(-)	GTGA	S000378
SEF4MOTIFGM7S	1002	(+)	RTTTTTR	S000103
GT1CONSENSUS	1005	(-)	GRWAAW	S000198
IBOXCORE	1006	(-)	GATAA	S000199
GATABOX	1007	(-)	GATA	S000039
ASF1MOTIFCAMV	1010	(-)	TGACG	S000024
WRKY71OS	1011	(-)	TGAC	S000447
GTGANTG10	1012	(-)	GTGA	S000378
CACTFTPPCA1	1017	(+)	YACT	S000449
CAATBOX1	1025	(-)	CAAT	S000028
WBOXATNPR1	1026	(+)	TTGAC	S000390
WRKY71OS	1027	(+)	TGAC	S000447
BIHD1OS	1027	(-)	TGTCA	S000498
CAATBOX1	1031	(-)	CAAT	S000028
GATABOX	1034	(+)	GATA	S000039
GT1CONSENSUS	1034	(+)	GRWAAW	S000198
IBOXCORE	1034	(+)	GATAA	S000199
CAATBOX1	1043	(+)	CAAT	S000028
ROOTMOTIFTAPOX1	1044	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	1045	(+)	ATATT	S000098
-10PEHVPSBD	1046	(+)	TATTCT	S000392
SORLIP1AT	1053	(+)	GCCAC	S000482
CAATBOX1	1057	(+)	CAAT	S000028
ARR1AT	1058	(-)	NGATT	S000454
GT1CONSENSUS	1064	(-)	GRWAAW	S000198
DOFCOREZM	1069	(-)	AAAG	S000265
TAAAGSTKST1	1069	(-)	TAAAG	S000387
NTBBF1ARROLB	1074	(+)	ACTTTA	S000273
DOFCOREZM	1075	(-)	AAAG	S000265
TAAAGSTKST1	1075	(-)	TAAAG	S000387
POLASIG1	1076	(-)	AATAAA	S000080
TATABOX5	1077	(+)	TTATTT	S000203
GT1CONSENSUS	1079	(-)	GRWAAW	S000198
ANAERO3CONSENSUS	1083	(+)	TCATCAC	S000479
GTGANTG10	1086	(-)	GTGA	S000378
AMMORESIIUDCRNIA1	1088	(-)	GGWAGGGT	S000374
CAREOSREP1	1098	(+)	CAACTC	S000421
MYBCORE	1109	(+)	CNGTTR	S000176
RAV1AAT	1110	(-)	CAACA	S000314
DOFCOREZM	1118	(+)	AAAG	S000265
ARR1AT	1121	(+)	NGATT	S000454
GT1CONSENSUS	1129	(-)	GRWAAW	S000198
IBOXCORE	1130	(-)	GATAA	S000199
GATABOX	1131	(-)	GATA	S000039
SEF4MOTIFGM7S	1139	(-)	RTTTTTR	S000103
ARR1AT	1143	(-)	NGATT	S000454
ROOTMOTIFTAPOX1	1148	(+)	ATATT	S000098
SEF1MOTIF	1148	(+)	ATATTTAWW	S000006
TATABOXOSPAL	1149	(+)	TATTTAA	S000400
POLASIG2	1151	(-)	AATTAAA	S000081
ERELEE4	1165	(-)	AWTTCAAA	S000037
CAATBOX1	1171	(-)	CAAT	S000028
CCAATBOX1	1171	(-)	CCAAT	S000030
TBOXATGAPB	1188	(+)	ACTTTG	S000383
-300ELEMENT	1188	(-)	TGHAAARK	S000122
DOFCOREZM	1189	(-)	AAAG	S000265
PROLAMINBOXOSGLUB1	1189	(-)	TGCAAAG	S000354
RYREPEATBNNAPA	1192	(-)	CATGCA	S000264
GT1CONSENSUS	1217	(-)	GRWAAW	S000198
S1FBOXSORPS1L21	1219	(-)	ATGGTA	S000223
ARR1AT	1238	(-)	NGATT	S000454
CAATBOX1	1256	(-)	CAAT	S000028
CACGCAATGMGH3	1256	(-)	CACGCAAT	S000368
GTGANTG10	1263	(+)	GTGA	S000378
WBOXHVISO1	1273	(+)	TGACT	S000442
WBOXNTERF3	1273	(+)	TGACY	S000457
WRKY71OS	1273	(+)	TGAC	S000447

ARR1AT	1286	(+)	NGATT	S000454
RAV1AAT	1297	(+)	CAACA	S000314
ARR1AT	1312	(-)	NGATT	S000454
GATABOX	1320	(-)	GATA	S000039
GT1CONSENSUS	1324	(+)	GRWAAW	S000198
GT1CONSENSUS	1325	(+)	GRWAAW	S000198
GT1GMSCAM4	1325	(+)	GAAAAA	S000453
CAATBOX1	1331	(+)	CAAT	S000028
CCAATBOX1	1336	(+)	CCAAT	S000030
CAATBOX1	1337	(+)	CAAT	S000028
REALPHALGLHCB21	1341	(-)	AACCAA	S000362
MYB1AT	1342	(-)	WAACCA	S000408
EBOXBNNAPA	1351	(+)	CANNTG	S000144
MYCCONSENSUSAT	1351	(+)	CANNTG	S000407
EBOXBNNAPA	1351	(-)	CANNTG	S000144
MYCCONSENSUSAT	1351	(-)	CANNTG	S000407
CACTFTPPCA1	1361	(+)	YACT	S000449
DOFCOREZM	1369	(-)	AAAG	S000265
ANAERO1CONSENSUS	1370	(-)	AAACAAA	S000477
CACTFTPPCA1	1388	(-)	YACT	S000449
ARR1AT	1391	(-)	NGATT	S000454
MRNA3ENDTAH3	1399	(-)	AATGGAAATG	S000069
GT1CONSENSUS	1400	(-)	GRWAAW	S000198
GT1CONSENSUS	1406	(-)	GRWAAW	S000198
GT1CONSENSUS	1407	(-)	GRWAAW	S000198
GTGANTG10	1418	(+)	GTGA	S000378
P1BS	1420	(+)	GNATATNC	S000459
P1BS	1420	(-)	GNATATNC	S000459
ROOTMOTIFTAPOX1	1421	(-)	ATATT	S000098
S1FBOXSORPS1L21	1429	(+)	ATGGTA	S000223
MYBPZM	1446	(+)	CCWACC	S000179
RAV1AAT	1468	(+)	CAACA	S000314
GT1CONSENSUS	1475	(-)	GRWAAW	S000198
CAATBOX1	1483	(+)	CAAT	S000028
POLASIG1	1484	(+)	AATAAA	S000080
SEF4MOTIFGM7S	1486	(-)	RTTTTTR	S000103
ARR1AT	1500	(-)	NGATT	S000454
GTGANTG10	1502	(-)	GTGA	S000378
TAAAGSTKST1	1521	(+)	TAAAG	S000387
NTBBF1ARROLB	1521	(-)	ACTTTA	S000273
DOFCOREZM	1522	(+)	AAAG	S000265
WBOXHVISO1	1524	(-)	TGACT	S000442
WBOXNTERF3	1524	(-)	TGACY	S000457
WRKY71OS	1525	(-)	TGAC	S000447
GTGANTG10	1526	(-)	GTGA	S000378
CACTFTPPCA1	1527	(+)	YACT	S000449
CCAATBOX1	1532	(+)	CCAAT	S000030
CAATBOX1	1533	(+)	CAAT	S000028
WBOXHVISO1	1536	(+)	TGACT	S000442
WBOXNTERF3	1536	(+)	TGACY	S000457
WRKY71OS	1536	(+)	TGAC	S000447
LTRECOREATCOR15	1543	(+)	CCGAC	S000153
GTGANTG10	1549	(-)	GTGA	S000378
DOFCOREZM	1559	(-)	AAAG	S000265
POLLEN1LELAT52	1560	(-)	AGAAA	S000245
DOFCOREZM	1563	(-)	AAAG	S000265
SEBFCONSSTPR10A	1565	(+)	YTGTCWC	S000391
ARFAT	1566	(+)	TGTCTC	S000270
SURECOREATSULTR11	1567	(-)	GAGAC	S000499
GT1CONSENSUS	1577	(-)	GRWAAW	S000198
ROOTMOTIFTAPOX1	1585	(-)	ATATT	S000098

Appendix 2B: PLACE motifs identified within the chalcone synthase promoter

Site Name	Location	Strand	Sequence	SITE ID
ROOTMOTIFTAPOX1	1	(+)	ATATT	S000098
SEF1MOTIF	1	(+)	ATATTTAWW	S000006
TATABOX2	3	(-)	TATAAAT	S000109
TATABOX5	10	(-)	TTATTT	S000203
POLASIG1	11	(+)	AATAAA	S000080
LECPLEACS2	13	(+)	TAAAATAT	S000465
ROOTMOTIFTAPOX1	16	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	19	(+)	ATATT	S000098
ROOTMOTIFTAPOX1	31	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	32	(+)	ATATT	S000098
TATABOX3	33	(+)	TATTAAT	S000110
TATABOX2	38	(-)	TATAAAT	S000109
DOFCOREZM	46	(-)	AAAG	S000265
ROOTMOTIFTAPOX1	57	(+)	ATATT	S000098
TATABOX4	60	(-)	TATATAA	S000111
TATABOX4	61	(+)	TATATAA	S000111
ROOTMOTIFTAPOX1	73	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	74	(+)	ATATT	S000098
CURECORECR	81	(+)	GTAC	S000493
CURECORECR	81	(-)	GTAC	S000493
CAATBOX1	85	(-)	CAAT	S000028
GATABOX	88	(+)	GATA	S000039
CACTFTPPCA1	91	(-)	YACT	S000449
GATABOX	95	(-)	GATA	S000039
NODCON1GM	96	(-)	AAAGAT	S000461
OSE1ROOTNODULE	96	(-)	AAAGAT	S000467
DOFCOREZM	98	(-)	AAAG	S000265
POLASIG1	102	(-)	AATAAA	S000080
CCA1ATLHCB1	113	(+)	AAMAATCT	S000149
CAATBOX1	115	(+)	CAAT	S000028
ARR1AT	116	(-)	NGATT	S000454
NODCON1GM	117	(-)	AAAGAT	S000461
OSE1ROOTNODULE	117	(-)	AAAGAT	S000467
DOFCOREZM	119	(-)	AAAG	S000265
TAAAGSTKST1	119	(-)	TAAAG	S000387
EBOXBNNAPA	133	(+)	CANNTG	S000144
MYCATERD1	133	(+)	CATGTG	S000413
MYCCONSENSUSAT	133	(+)	CANNTG	S000407
DPBFCOREDCDC3	133	(-)	ACACNNG	S000292
EBOXBNNAPA	133	(-)	CANNTG	S000144
MYCATRD22	133	(-)	CACATG	S000174
MYCCONSENSUSAT	133	(-)	CANNTG	S000407
DOFCOREZM	140	(-)	AAAG	S000265
POLLEN1LELAT52	141	(-)	AGAAA	S000245
DOFCOREZM	144	(-)	AAAG	S000265
ROOTMOTIFTAPOX1	153	(-)	ATATT	S000098
GATABOX	155	(-)	GATA	S000039
SEF4MOTIFGM7S	166	(-)	RTTTTTR	S000103
ROOTMOTIFTAPOX1	170	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	171	(+)	ATATT	S000098
LECPLEACS2	171	(-)	TAAAATAT	S000465
GATABOX	180	(+)	GATA	S000039
GATABOX	187	(+)	GATA	S000039
CACTFTPPCA1	189	(+)	YACT	S000449
CACTFTPPCA1	209	(+)	YACT	S000449
TATABOX4	212	(-)	TATATAA	S000111
ROOTMOTIFTAPOX1	218	(+)	ATATT	S000098
SEF4MOTIFGM7S	220	(+)	RTTTTTR	S000103
ROOTMOTIFTAPOX1	226	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	227	(+)	ATATT	S000098
TATABOXOSPAL	228	(+)	TATTTAA	S000400
DOFCOREZM	250	(-)	AAAG	S000265
GTGANTG10	255	(+)	GTGA	S000378
AACACOREOSGLUB1	258	(+)	AACAAAC	S000353
CPBCSPOR	270	(-)	TATTAG	S000491
ROOTMOTIFTAPOX1	272	(-)	ATATT	S000098
GATABOX	274	(-)	GATA	S000039

RAV1AAT	281	(+)	CAACA	S000314
SEF4MOTIFGM7S	285	(+)	RTTTTTR	S000103
WRKY71OS	291	(-)	TGAC	S000447
GTGANTG10	305	(-)	GTGA	S000378
CACTFTPPCA1	308	(+)	YACT	S000449
NODCON2GM	310	(+)	CTCTT	S000462
OSE2ROOTNODULE	310	(+)	CTCTT	S000468
DOFCOREZM	312	(-)	AAAG	S000265
ACGTATERD1	318	(+)	ACGT	S000415
HEXMOTIFTAH3H4	318	(+)	ACGTCA	S000053
ACGTATERD1	318	(-)	ACGT	S000415
TGACGTVMAMY	318	(-)	TGACGT	S000377
ASF1MOTIFCAMV	319	(-)	TGACG	S000024
WRKY71OS	320	(-)	TGAC	S000447
LECPLEACS2	334	(+)	TAAAATAT	S000465
ROOTMOTIFTAPOX1	337	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	340	(+)	ATATT	S000098
DOFCOREZM	354	(+)	AAAG	S000265
-10PEHVPSBD	356	(-)	TATTCT	S000392
CACTFTPPCA1	360	(+)	YACT	S000449
DOFCOREZM	362	(-)	AAAG	S000265
ARR1AT	369	(-)	NGATT	S000454
NODCON1GM	370	(-)	AAAGAT	S000461
OSE1ROOTNODULE	370	(-)	AAAGAT	S000467
DOFCOREZM	372	(-)	AAAG	S000265
TAAAGSTKST1	372	(-)	TAAAG	S000387
CACTFTPPCA1	375	(+)	YACT	S000449
WBOXATNPR1	378	(+)	TTGAC	S000390
WBOXHVISO1	379	(+)	TGACT	S000442
WBOXNTERF3	379	(+)	TGACY	S000457
WRKY71OS	379	(+)	TGAC	S000447
DOFCOREZM	382	(-)	AAAG	S000265
AACACOREOSGLUB1	390	(-)	AACAAAC	S000353
ANAERO1CONSENSUS	391	(-)	AAACAAA	S000477
GTGANTG10	397	(-)	GTGA	S000378
LECPLEACS2	405	(+)	TAAAATAT	S000465
ROOTMOTIFTAPOX1	408	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	411	(+)	ATATT	S000098
SEF1MOTIF	411	(+)	ATATTTAWW	S000006
TATABOXOSPAL	412	(+)	TATTTAA	S000400
DOFCOREZM	425	(+)	AAAG	S000265
ROOTMOTIFTAPOX1	439	(-)	ATATT	S000098
EBOXBNNAPA	449	(+)	CANNTG	S000144
MYCCONSENSUSAT	449	(+)	CANNTG	S000407
EBOXBNNAPA	449	(-)	CANNTG	S000144
MYCCONSENSUSAT	449	(-)	CANNTG	S000407
T/GBOXATPIN2	461	(+)	AACGTG	S000458
ABRELATERD1	462	(+)	ACGTG	S000414
ACGTATERD1	462	(+)	ACGT	S000415
ACGTATERD1	462	(-)	ACGT	S000415
CURECORECR	467	(+)	GTAC	S000493
CURECORECR	467	(-)	GTAC	S000493
CAATBOX1	470	(+)	CAAT	S000028
ARR1AT	471	(-)	NGATT	S000454
CACTFTPPCA1	482	(-)	YACT	S000449
ROOTMOTIFTAPOX1	485	(+)	ATATT	S000098
CAATBOX1	487	(-)	CAAT	S000028
GT1CONSENSUS	490	(+)	GRWAAW	S000198
GT1GMSCAM4	490	(+)	GAAAAA	S000453
DOFCOREZM	494	(+)	AAAG	S000265
CACTFTPPCA1	513	(-)	YACT	S000449
GTGANTG10	514	(+)	GTGA	S000378
REALPHALGLHCB21	519	(+)	AACCAA	S000362
RAV1AAT	522	(+)	CAACA	S000314
MYBCORE	522	(-)	CNGTTR	S000176
POLLEN1LELAT52	526	(+)	AGAAA	S000245
SITEIIATCYTC	550	(+)	TGGGCY	S000474
CACGTGMOTIF	560	(+)	CACGTG	S000042
EBOXBNNAPA	560	(+)	CANNTG	S000144
MYCCONSENSUSAT	560	(+)	CANNTG	S000407
ABRELATERD1	560	(-)	ACGTG	S000414
CACGTGMOTIF	560	(-)	CACGTG	S000042
EBOXBNNAPA	560	(-)	CANNTG	S000144

MYCCONSENSUSAT	560	(-)	CANNTG	S000407
ABRELATERD1	561	(+)	ACGTG	S000414
ACGTATERD1	561	(+)	ACGT	S000415
ACGTATERD1	561	(-)	ACGT	S000415
GTGANTG10	563	(+)	GTGA	S000378
ARR1AT	566	(-)	NGATT	S000454
GATABOX	595	(-)	GATA	S000039
TATABOX5	599	(+)	TTATTT	S000203
TATABOX2	601	(-)	TATAAAT	S000109
WBOXATNPR1	625	(-)	TTGAC	S000390
WRKY71OS	625	(-)	TGAC	S000447
RAV1AAT	627	(+)	CAACA	S000314
PYRIMIDINEBOXOSRAMY1A	635	(+)	CCTTTT	S000259
DOFCOREZM	636	(-)	AAAG	S000265
PYRIMIDINEBOXOSRAMY1A	644	(+)	CCTTTT	S000259
DOFCOREZM	645	(-)	AAAG	S000265
GT1CONSENSUS	648	(-)	GRWAAW	S000198
S1FBOXSORPS1L21	650	(-)	ATGGTA	S000223
CAREOSREP1	677	(+)	CAACTC	S000421
MYBCORE	694	(+)	CNGTTR	S000176
MYB2CONSENSUSAT	694	(-)	YAACKG	S000409
SEF4MOTIFGM7S	703	(+)	RTTTTTR	S000103
CURECORECR	709	(+)	GTAC	S000493
CURECORECR	709	(-)	GTAC	S000493
LTRE1HVBLT49	715	(+)	CCGAAA	S000250
GT1CONSENSUS	717	(+)	GRWAAW	S000198
GT1GMSCAM4	717	(+)	GAAAAA	S000453
DOFCOREZM	720	(+)	AAAG	S000265

Appendix 2C: PLACE motifs identified within the granule-bound starch synthase promoter

Site Name	Location	Strand	Sequence	SITE ID
ARR1AT	5	(-)	NGATT	S000454
ARR1AT	8	(+)	NGATT	S000454
CAATBOX1	22	(-)	CAAT	S000028
GT1CONSENSUS	35	(+)	GRWAAW	S000198
TATABOX5	37	(-)	TTATTT	S000203
CAATBOX1	47	(-)	CAAT	S000028
ARR1AT	49	(+)	NGATT	S000454
ERELEE4	60	(+)	AWTTCAAA	S000037
TATABOX5	66	(-)	TTATTT	S000203
POLASIG3	67	(+)	AATAAT	S000088
SEF4MOTIFGM7S	71	(+)	RTTTTTR	S000103
TAAAGSTKST1	76	(+)	TAAAG	S000387
DOFCOREZM	77	(+)	AAAG	S000265
NODCON1GM	77	(+)	AAAGAT	S000461
OSE1ROOTNODULE	77	(+)	AAAGAT	S000467
AGMOTIFNTMYB2	79	(+)	AGATCCAA	S000444
ROOTMOTIFTAPOX1	86	(-)	ATATT	S000098
ROOTMOTIFTAPOX1	87	(+)	ATATT	S000098
SEF1MOTIF	87	(+)	ATATTTAWW	S000006
TATABOX2	89	(-)	TATAAAT	S000109
ARR1AT	97	(-)	NGATT	S000454
TATABOX2	107	(-)	TATAAAT	S000109
TATAPVTRNALEU	108	(+)	TTTATATA	S000340
TATABOX4	109	(-)	TATATAA	S000111
ROOTMOTIFTAPOX1	113	(+)	ATATT	S000098
SEF4MOTIFGM7S	115	(+)	RTTTTTR	S000103
SEF4MOTIFGM7S	120	(-)	RTTTTTR	S000103
MARTBOX	133	(-)	TTWTWTTWTT	S000067
TATABOX5	135	(-)	TTATTT	S000203
POLASIG1	136	(+)	AATAAA	S000080
POLLEN1LELAT52	145	(-)	AGAAA	S000245
TAAAGSTKST1	149	(+)	TAAAG	S000387
DOFCOREZM	150	(+)	AAAG	S000265
-10PEHVPSBD	152	(-)	TATTCT	S000392
ROOTMOTIFTAPOX1	154	(-)	ATATT	S000098
TATABOX4	156	(+)	TATATAA	S000111
TATAPVTRNALEU	156	(-)	TTTATATA	S000340
DOFCOREZM	167	(+)	AAAG	S000265
PYRIMIDINEBOXOSRAMY1A	173	(-)	CCTTTT	S000259
DOFCOREZM	174	(+)	AAAG	S000265
ARR1AT	178	(+)	NGATT	S000454
CAATBOX1	189	(+)	CAAT	S000028
POLASIG3	190	(+)	AATAAT	S000088
ARR1AT	200	(+)	NGATT	S000454
TATABOXOSPAL	204	(-)	TATTTAA	S000400
TATABOX5	206	(-)	TTATTT	S000203
CACTFTPPCA1	211	(-)	YACT	S000449
ARR1AT	221	(+)	NGATT	S000454
GT1CONSENSUS	230	(-)	GRWAAW	S000198
IBOXCORE	231	(-)	GATAA	S000199
GATABOX	232	(-)	GATA	S000039
CACTFTPPCA1	237	(+)	YACT	S000449
GATABOX	245	(-)	GATA	S000039
MYBST1	245	(-)	GGATA	S000180
PYRIMIDINEBOXOSRAMY1A	248	(+)	CCTTTT	S000259
DOFCOREZM	249	(-)	AAAG	S000265
GT1CONSENSUS	252	(-)	GRWAAW	S000198
GT1GMSCAM4	252	(-)	GAAAAA	S000453
POLLEN1LELAT52	254	(-)	AGAAA	S000245
NODCON2GM	257	(+)	CTCTT	S000462
OSE2ROOTNODULE	257	(+)	CTCTT	S000468
TATABOX5	260	(+)	TTATTT	S000203
DOFCOREZM	268	(-)	AAAG	S000265
ERELEE4	269	(-)	AWTTCAAA	S000037
POLASIG2	274	(+)	AATTAAA	S000081
ARR1AT	285	(-)	NGATT	S000454
ROOTMOTIFTAPOX1	290	(-)	ATATT	S000098

MYB1AT	298	(-)	WAACCA	S000408
TATABOX2	307	(-)	TATAAAT	S000109
TATAPVTRNALEU	308	(+)	TTTATATA	S000340
TATABOX4	309	(-)	TATATAA	S000111
CACTFTPPCA1	314	(+)	YACT	S000449
BOXIINTPATPB	323	(+)	ATAGAA	S000296
S1FBOXSORPS1L21	337	(+)	ATGGTA	S000223
WBOXATNPR1	356	(+)	TTGAC	S000390
WRKY71OS	357	(+)	TGAC	S000447
POLASIG2	364	(-)	AATTAAA	S000081
ARR1AT	376	(-)	NGATT	S000454
REALPHALGLHCB21	386	(+)	AACCAA	S000362
CCAATBOX1	388	(+)	CCAAT	S000030
CAATBOX1	389	(+)	CAAT	S000028
ARR1AT	390	(-)	NGATT	S000454
CACTFTPPCA1	395	(+)	YACT	S000449
INRNTPSADB	397	(+)	YTCANTYY	S000395
CAATBOX1	399	(+)	CAAT	S000028
INRNTPSADB	402	(+)	YTCANTYY	S000395
CAATBOX1	404	(+)	CAAT	S000028
ARR1AT	409	(+)	NGATT	S000454
CAREOSREP1	414	(+)	CAACTC	S000421
POLASIG2	426	(+)	AATTAAA	S000081
GTGANTG10	440	(-)	GTGA	S000378
MYBPZM	444	(+)	CCWACC	S000179
ERELEE4	451	(+)	AWTTCAAA	S000037
POLASIG2	461	(+)	AATTAAA	S000081
POLASIG2	466	(+)	AATTAAA	S000081
POLASIG2	471	(+)	AATTAAA	S000081
POLASIG2	476	(+)	AATTAAA	S000081
RAV1AAT	484	(+)	CAACA	S000314
TATABOX5	489	(-)	TTATTT	S000203
POLASIG1	490	(+)	AATAAA	S000080
CAREOSREP1	502	(+)	CAACTC	S000421
CAATBOX1	507	(+)	CAAT	S000028
POLASIG1	508	(+)	AATAAA	S000080
-10PEHVPSBD	515	(+)	TATTCT	S000392
RBCSCONSENSUS	533	(+)	AATCCAA	S000127
ARR1AT	533	(-)	NGATT	S000454
DOFCOREZM	555	(+)	AAAG	S000265
ARR1AT	571	(+)	NGATT	S000454
SEF4MOTIFGM7S	573	(+)	RTTTTTR	S000103
CACTFTPPCA1	580	(+)	YACT	S000449
SEF4MOTIFGM7S	583	(-)	RTTTTTR	S000103
NODCON2GM	590	(-)	CTCTT	S000462
OSE2ROOTNODULE	590	(-)	CTCTT	S000468
CACTFTPPCA1	593	(-)	YACT	S000449
PYRIMIDINEBOXOSRAMY1A	599	(-)	CCTTTT	S000259
DOFCOREZM	600	(+)	AAAG	S000265
SEF4MOTIFGM7S	608	(-)	RTTTTTR	S000103
SORLIP1AT	618	(+)	GCCAC	S000482
ASF1MOTIFCAMV	637	(-)	TGACG	S000024
WBOXATNPR1	638	(-)	TTGAC	S000390
WRKY71OS	638	(-)	TGAC	S000447
CAATBOX1	640	(+)	CAAT	S000028
INRNTPSADB	648	(+)	YTCANTYY	S000395
CAATBOX1	650	(+)	CAAT	S000028
CACTFTPPCA1	658	(-)	YACT	S000449
CURECORECR	659	(+)	GTAC	S000493
CURECORECR	659	(-)	GTAC	S000493
CACTFTPPCA1	660	(+)	YACT	S000449
GATABOX	674	(+)	GATA	S000039
GATABOX	676	(-)	GATA	S000039
GT1CONSENSUS	681	(+)	GRWAAW	S000198
DOFCOREZM	684	(+)	AAAG	S000265
CAATBOX1	688	(+)	CAAT	S000028
SV40COREENHAN	688	(-)	GTGGWWHG	S000123
ARR1AT	689	(-)	NGATT	S000454
DOFCOREZM	699	(-)	AAAG	S000265
GT1CONSENSUS	709	(-)	GRWAAW	S000198
GT1GMSCAM4	709	(-)	GAAAAA	S000453
GTGANTG10	713	(-)	GTGA	S000378
EBOXBNNAPA	728	(+)	CANNTG	S000144

MYCCONSENSUSAT	728	(+)	CANNTG	S000407
EBOXBNNAPA	728	(-)	CANNTG	S000144
MYCCONSENSUSAT	728	(-)	CANNTG	S000407
RAV1AAT	732	(-)	CAACA	S000314
SITEIIATCYTC	738	(-)	TGGGCY	S000474
MYBCORE	745	(+)	CNGTTR	S000176
MYB2AT	745	(-)	TAACTG	S000177
MYB2CONSENSUSAT	745	(-)	YAACKG	S000409
SEF3MOTIFGM	753	(+)	AACCCA	S000115
INRNTPSADB	763	(-)	YTCANTYY	S000395
GTGANTG10	773	(+)	GTGA	S000378
WBOXHVISO1	774	(+)	TGACT	S000442
WBOXNTERF3	774	(+)	TGACY	S000457
WRKY71OS	774	(+)	TGAC	S000447
CACTFTPPCA1	779	(-)	YACT	S000449
GTGANTG10	780	(+)	GTGA	S000378
WBOXNTERF3	781	(+)	TGACY	S000457
WRKY71OS	781	(+)	TGAC	S000447
GT1CONSENSUS	790	(-)	GRWAAW	S000198
INTRONLOWER	794	(-)	TGCAGG	S000086
CACTFTPPCA1	800	(+)	YACT	S000449
PREATPRODH	801	(+)	ACTCAT	S000450
S1FBOXSORPS1L21	806	(-)	ATGGTA	S000223
SORLIP2AT	817	(+)	GGGCC	S000483
SORLIP2AT	818	(-)	GGGCC	S000483
BOXLCOREDCPAL	825	(+)	ACCWWCC	S000492
ABREZMRAB28	830	(+)	CCACGTGG	S000133
ABREZMRAB28	830	(-)	CCACGTGG	S000133
CACGTGMOTIF	831	(+)	CACGTG	S000042
EBOXBNNAPA	831	(+)	CANNTG	S000144
MYCCONSENSUSAT	831	(+)	CANNTG	S000407
ABRELATERD1	831	(-)	ACGTG	S000414
CACGTGMOTIF	831	(-)	CACGTG	S000042
EBOXBNNAPA	831	(-)	CANNTG	S000144
MYCCONSENSUSAT	831	(-)	CANNTG	S000407
ABRELATERD1	832	(+)	ACGTG	S000414
ACGTATERD1	832	(+)	ACGT	S000415
ACGTATERD1	832	(-)	ACGT	S000415
CACTFTPPCA1	842	(+)	YACT	S000449
CAATBOX1	851	(-)	CAAT	S000028
UPRMOTIFIIAT	868	(+)	CCNNNNNNNNNNNNCCACG	S000426
TAAAGSTKST1	877	(+)	TAAAG	S000387
DOFCOREZM	878	(+)	AAAG	S000265
SORLIP1AT	881	(+)	GCCAC	S000482
ACGTABREMOTIFA2OSEM	881	(-)	ACGTGKC	S000394
BOXIIPCCHS	881	(-)	ACGTGGC	S000229
ABRELATERD1	883	(-)	ACGTG	S000414
T/GBOXATPIN2	883	(-)	AACGTG	S000458
ACGTATERD1	884	(+)	ACGT	S000415
ACGTATERD1	884	(-)	ACGT	S000415
SORLREP2AT	884	(-)	ATAAAACGT	S000487
POLASIG1	888	(-)	AATAAA	S000080
CAATBOX1	891	(-)	CAAT	S000028
ACGTTBOX	901	(+)	AACGTT	S000132
ACGTTBOX	901	(-)	AACGTT	S000132
ACGTATERD1	902	(+)	ACGT	S000415
ACGTATERD1	902	(-)	ACGT	S000415
ARR1AT	909	(-)	NGATT	S000454
CAATBOX1	920	(-)	CAAT	S000028
CCAATBOX1	920	(-)	CCAAT	S000030
GTGANTG10	935	(+)	GTGA	S000378
ARE1	944	(+)	RGTGACNNNGC	S000022
CACTFTPPCA1	944	(-)	YACT	S000449
GTGANTG10	945	(+)	GTGA	S000378
WBOXNTERF3	946	(+)	TGACY	S000457
WRKY71OS	946	(+)	TGAC	S000447
SORLIP1AT	953	(+)	GCCAC	S000482
CACTFTPPCA1	955	(+)	YACT	S000449
RAV1AAT	960	(+)	CAACA	S000314
GTGANTG10	972	(-)	GTGA	S000378
CACTFTPPCA1	973	(+)	YACT	S000449
SEF3MOTIFGM	980	(+)	AACCCA	S000115
EBOXBNNAPA	984	(+)	CANNTG	S000144

MYCCONSENSUSAT	984	(+)	CANNTG	S000407
EBOXBNNAPA	984	(-)	CANNTG	S000144
MYCCONSENSUSAT	984	(-)	CANNTG	S000407
CACTFTPPCA1	990	(+)	YACT	S000449
CACTFTPPCA1	995	(-)	YACT	S000449
GTGANTG10	996	(+)	GTGA	S000378
CACTFTPPCA1	999	(-)	YACT	S000449
GTGANTG10	1000	(+)	GTGA	S000378
CURECORECR	1023	(+)	GTAC	S000493
CURECORECR	1023	(-)	GTAC	S000493
CACTFTPPCA1	1024	(+)	YACT	S000449
ARR1AT	1039	(+)	NGATT	S000454
CACTFTPPCA1	1047	(-)	YACT	S000449
GTGANTG10	1048	(+)	GTGA	S000378
CACTFTPPCA1	1054	(+)	YACT	S000449
GATABOX	1059	(-)	GATA	S000039
MYBCORE	1062	(+)	CNGTTR	S000176
MYB2AT	1062	(-)	TAACTG	S000177
MYB2CONSENSUSAT	1062	(-)	YAACKG	S000409
PYRIMIDINEBOXOSRAMY1A	1083	(+)	CCTTTT	S000259
-300ELEMENT	1083	(-)	TGHAAARK	S000122
DOFCOREZM	1084	(-)	AAAG	S000265
ANAERO3CONSENSUS	1088	(+)	TCATCAC	S000479
GTGANTG10	1091	(-)	GTGA	S000378
ACGTATERD1	1098	(+)	ACGT	S000415
ACGTATERD1	1098	(-)	ACGT	S000415
NODCON2GM	1103	(+)	CTCTT	S000462
OSE2ROOTNODULE	1103	(+)	CTCTT	S000468
TATABOX4	1113	(+)	TATATAA	S000111
TATAPVTRNALEU	1113	(-)	TTTATATA	S000340
RAV1AAT	1122	(-)	CAACA	S000314
SEF3MOTIFGM	1125	(-)	AACCCA	S000115
SEF4MOTIFGM7S	1128	(+)	RTTTTTR	S000103
DOFCOREZM	1136	(-)	AAAG	S000265
TAAAGSTKST1	1136	(-)	TAAAG	S000387
REALPHALGLHCB21	1147	(-)	AACCAA	S000362
MYB1AT	1148	(-)	WAACCA	S000408
MYBPLANT	1150	(-)	MACCWAMC	S000167
REALPHALGLHCB21	1152	(-)	AACCAA	S000362
MYB1AT	1153	(-)	WAACCA	S000408
POLLEN1LELAT52	1157	(-)	AGAAA	S000245
DOFCOREZM	1160	(-)	AAAG	S000265
GATABOX	1168	(-)	GATA	S000039
DOFCOREZM	1174	(-)	AAAG	S000265
CAATBOX1	1182	(+)	CAAT	S000028

Bibliography

Adamczyk, J., D. Hardee, L. Adams, and D. Sumerford. 2001a. Correlating differences in larval survival and development of bollworm (Lepidoptera: Noctuidae) and fall armyworm (Lepidoptera: Noctuidae) to differential expression of Cry1A(c) delta-endotoxin in various plant parts among commercial cultivars of transgenic Bacillus thuringiensis cotton. J Econ Entomol 94: 284-90.

Adamczyk, J. J., and D. Sumerford. 2001. Potential factors impacting season-long expression of Cry1Ac in 13 commercial varieties of Bollgard cotton. J Insect Sci 1: 13.

Adamczyk, J. J., L. C. Adams, and D. D. Hardee. 2001b. Field efficacy and seasonal expression profiles for terminal leaves of single and double Bacillus thuringiensis toxin cotton genotypes. Journal of Economic Entomology 94: 1589-1593.

Agius, F., I. Amaya, M. A. Botella, and V. Valpuesta. 2005. Functional analysis of homologous and heterologous promoters in strawberry fruits using transient expression. Journal of Experimental Botany 56: 37-46.

Aharoni, A., A. P. Giri, F. W. A. Verstappen, C. M. Bertea, R. Sevenier, Z. K. Sun, M. A. Jongsma, W. Schwab, and H. J. Bouwmeester. 2004. Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. Plant Cell 16: 3110-3131.

Alcantara, E. P., R. M. Aguda, A. Curtiss, D. H. Dean, and M. B. Cohen. 2004. Bacillus thuringiensis delta-endotoxin binding to brush border membrane vesicles of rice stem borers. Archives of Insect Biochemistry and Physiology 55: 169-177.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 215: 403-410.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-402.

Arioli, T., P. A. Howles, J. J. Weinman, and B. G. Rolfe. 1994. In Trifolium-Subterraneum, Chalcone Synthase Is Encoded by a Multigene Family. Gene 138: 79-86.

Babu, R. M., A. Sajeena, K. Seetharaman, and M. S. Reddy. 2003. Advances in genetically engineered (transgenic) plants in pest management - an over view. Crop Protection 22: 1071-1086.

Baker, S. S., C. L. Rugh, and J. C. Kamalay. 1990. Rna and DNA Isolation from Recalcitrant Plant-Tissues. Biotechniques 9: 268-&.

Baldwin, T. C., A. J. van Hengel, and K. Roberts. 2000. The C-terminal PAC domain of a secreted arabinogalactan-protein from carrot defines a family of basic proline-rich proteins, pp. 43–50. *In* E. A. B. Nothnagel, A. and Clarke, A.E. [ed.], Cell and Developmental Biology of Arabinogalactan-Proteins. Kluwer/Plenum, New York.

Baldwin, T. C., C. Domingo, T. Schindler, G. Seetharaman, N. Stacey, and K. Roberts. 2001. DcAGP1, a secreted arabinogalactan protein, is related to a family of basic proline-rich proteins. Plant Molecular Biology 45: 421-435.

Ball, S. G., M. van de Wal, and R. G. F. Visser. 1998. Progress in understanding the biosynthesis of amylose. Trends in Plant Science 3: 462-467.

Basra, A. S., and C. P. Malik. 1984. Development of the cotton fibre. Int Rev Cytol 89: 65-112.

Baud, S., M. N. Vaultier, and C. Rochat. 2004. Structure and expression profile of the sucrose synthase multigene family in Arabidopsis. Journal of Experimental Botany 55: 397-409.

Baum, K., B. Groning, and I. Meier. 1997. Improved ballistic transient transformation conditions for tomato fruit allow identification of organ-specific contributions of I-box and G-box to the RBCS2 promoter activity. Plant Journal 12: 463-469.

Bednarz, C. W., and P. M. Roberts. 2001. Spatial yield distribution in cotton following early-season floral bud removal. Crop Science 41: 1800-1808.

Bender, J. 2004. Chromatin-based silencing mechanisms. Current Opinion in Plant Biology 7: 521-526.

Bent, A. F., and I. C. Yu. 1999. Applications of molecular biology to plant disease and insect resistance, pp. 251-298, Advances in Agronomy, Vol 66.

Berk, A. J. 1999. Activation of RNA polymerase II transcription. Curr Opin Cell Biol 11: 330-335.

Berliner, E. 1915. Über die schlaffsucht der mehlmottenraupe (Ephestia kuhniella, Zell.) und ihren erreger Bacillus thuringiensis n. sp. Z. Agnew. Entomol.: 29–35.

Bhattacharyya, S., S. Pattanaik, and I. B. Maiti. 2003. Intron-mediated enhancement of gene expression in transgenic plants using chimeric constructs composed of the Peanut chlorotic streak virus (PClSV) promoter-leader and the antisense orientation of PClSVORF VII (p7R). Planta 218: 115-124.

Bogs, J., M. O. Downey, J. S. Harvey, A. R. Ashton, G. J. Tanner, and S. P. Robinson. 2005. Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. Plant Physiology 139: 652-663.

Bondada, B. R., D. M. Oosterhuis, S. D. Wullschleger, K. S. Him, and W. M. Harris. 1994. Anatomical Considerations Related to Photosynthesis in Cotton (Gossypium-Hirsutum L) Leaves, Bracts, and the Capsule Wall. Journal of Experimental Botany 45: 111-118.

Boulter, D. 1993. Insect Pest-Control by Copying Nature Using Genetically-Engineered Crops. Phytochemistry 34: 1453-1466.

Brook, K. D., A. B. Hearn, and C. F. Kelly. 1992. Response of Cotton to Damage by Insect Pests in Australia - Pest-Management Trials. Journal of Economic Entomology 85: 1356-1367.

Buleon, A., P. Colonna, V. Planchot, and S. Ball. 1998. Starch granules: structure and biosynthesis. International Journal of Biological Macromolecules 23: 85-112.

Callis, J., M. Fromm, and V. Walbot. 1987. Introns Increase Gene-Expression in Cultured Maize Cells. Genes & Development 1: 1183-1200.

Carlini, C. R., and M. F. Grossi-de-Sa. 2002. Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides. Toxicon 40: 1515-1539.

Carriere, Y., C. Ellers-Kirk, Y. B. Liu, M. A. Sims, A. L. Patin, T. J. Dennehy, and B. E. Tabashnik. 2001a. Fitness costs and maternal effects associated with resistance to transgenic cotton in the pink bollworm (Lepidoptera : Gelechiidae). Journal of Economic Entomology 94: 1571-1576.

Carriere, Y., C. Ellers-Kirk, A. L. Patin, M. A. Sims, S. Meyer, Y. B. Liu, T. J. Dennehy, and B. E. Tabashnik. 2001b. Overwintering cost associated with resistance to transgenic cotton in the pink bollworm (Lepidoptera : Gelechiidae). Journal of Economic Entomology 94: 935-941.

Castaldini, M., A. Turrini, C. Sbrana, A. Benedetti, M. Marchionni, S. Mocali, A. Fabiani, S. Landi, F. Santomassimo, B. Pietrangeli, M. P. Nuti, N. Miclaus, and M. Giovannetti. 2005. Impact of Bt corn on rhizospheric and on beneficial mycorrhizal symbiosis and soil eubacterial communities iosis in experimental microcosms. Applied and Environmental Microbiology 71: 6719-6729.

Cedroni, M. L., R. C. Cronn, K. L. Adams, T. A. Wilkins, and J. F. Wendel. 2003. Evolution and expression of MYB genes in diploid and polyploid cotton. Plant Molecular Biology 51: 313-325.

Chandler, V. L., and H. Vaucheret. 2001. Gene activation and gene silencing. Plant Physiol 125: 145-8.

Chen, B., M. L. Colgrave, N. L. Daly, K. J. Rosengren, K. R. Gustafson, and D. J. Craik. 2005. Isolation and characterization of novel cyclotides from Viola hederaceae - Solution structure and anti-HIV activity of vhl-1, a leaf-specific expressed cyclotide. Journal of Biological Chemistry 280: 22395-22405.

Chen, C. G., S. L. Mau, and A. E. Clarke. 1993. Nucleotide-Sequence and Style-Specific Expression of a Novel Proline-Rich Protein Gene from Nicotiana-Alata. Plant Molecular Biology 21: 391-395.

Chopra, S., J. Delfavero, R. Dolferus, and M. Jacobs. 1992. Sucrose Synthase of Arabidopsis - Genomic Cloning and Sequence Characterization. Plant Molecular Biology 18: 131-134.

Claudot, A. C., D. Ernst, H. Sandermann, and A. Drouet. 1999. Cloning and characterization of two members of the chalcone synthase gene family from walnut. Plant Physiology and Biochemistry 37: 721-730.

Cline, J., J. C. Braman, and H. H. Hogrefe. 1996. PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Research 24: 3546-3551.

Crickmore, N., D. R. Zeigler, J. Feitelson, E. Schnepf, J. Van Rie, D. Lereclus, J. Baum, and D. H. Dean. 1998. Revision of the nomenclature for the Bacillus thuringiensis pesticidal crystal proteins. Microbiology and Molecular Biology Reviews 62: 807-+.

Curtis, M. D., and U. Grossniklaus. 2003. A Gateway cloning vector set for high-throughput analysis of genes *in planta*. Plant Physiology 133: 462-469.

Daniels, M. J., F. Chaumont, T. E. Mirkov, and M. J. Chrispeels. 1996. Characterization of a new vacuolar membrane aquaporin sensitive to mercury at a unique site. Plant Cell 8: 587-599.

D'Aoust, M. A., S. Yelle, and B. Nguyen-Quoc. 1999. Antisense inhibition of tomato fruit sucrose synthase decreases fruit setting and the sucrose unloading capacity of young fruit. Plant Cell 11: 2407-2418.

Datta, K., A. Vasquez, J. Tu, L. Torrizo, M. F. Alam, N. Oliva, E. Abrigo, G. S. Khush, and S. K. Datta. 1998. Constitutive and tissue-specific differential expression of the cryIA(b) gene in transgenic rice plants conferring resistance to rice insect pest. Theoretical and Applied Genetics 97: 20-30.

de Maagd, R. A., D. Bosch, and W. Stiekema. 1999. Bacillus thuringiensis toxin-mediated insect resistance in plants. Trends in Plant Science 4: 9-13.

Dejardin, A., C. Rochat, S. Wuilleme, and J. P. Boutin. 1997. Contribution of sucrose synthase, ADP-glucose pyrophosphorylase and starch synthase to starch synthesis in developing pea seeds. Plant Cell and Environment 20: 1421-1430.

Delaney, S. K. 2005. Molecular regulation of fibre-specific gene expression in cotton (gossypium hirsutum), School of Molecular and Biomedical Science (Discipline of Genetics). University of Adelaide, Adelaide.

Dellaporta, S. L., J. Wood, and J. B. Hicks. 1985. Maize DNA miniprep, pp. 36-37, Molecular Biology of Plants: A Laboratory Course Manual. Cold Spring Harbour Laboratory, New York.

Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387-395.

Devic, M., J. Guilleminot, I. Debeaujon, N. Bechtold, E. Bensaude, M. Koornneef, G. Pelletier, and M. Delseny. 1999. The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. Plant Journal 19: 387-398.

Dixon, R. A., D. Y. Xie, and S. B. Sharma. 2005. Proanthocyanidins - a final frontier in flavonoid research? New Phytologist 165: 9-28.

Endrizzi, J. E., E. L. Turcotte, and R. J. Kohel. 1985. Genetics, cytology and evolution of *Gossypium*. Adv Genet 23: 271-375.

Engstrom, E. M., A. Izhaki, and J. L. Bowman. 2004. Promoter bashing, microRNAs, and knox genes. New insights, regulators,, and targets-of-regulation in the establishment of lateral organ polarity in Arabidopsis. Plant Physiology 135: 685-694.

Estruch, J. J., S. Kadwell, E. Merlin, and L. Crossland. 1994. Cloning and Characterization of a Maize Pollen-Specific Calcium-Dependent Calmodulin-Independent Protein-Kinase. Proceedings of the National Academy of Sciences of the United States of America 91: 8837-8841.

Estruch, J. J., N. B. Carozzi, N. Desai, N. B. Duck, G. W. Warren, and M. G. Koziel. 1997. Transgenic plants: An emerging approach to pest control. Nature Biotechnology 15: 137-141.

Faktor, O., J. M. Kooter, R. A. Dixon, and C. J. Lamb. 1996. Functional dissection of a bean chalcone synthase gene promoter in transgenic tobacco plants reveals sequence motifs essential for floral expression. Plant Molecular Biology 32: 849-859.

Faktor, O., G. Loake, R. A. Dixon, and C. J. Lamb. 1997. The G-box and H-box in a 39 bp region of a French bean chalcone synthase promoter constitute a tissue-specific regulatory element. Plant Journal 11: 1105-1113.

Farrow, R. A., and J. C. Daly. 1987. Long-Range Movements as an Adaptive Strategy in the Genus Heliothis (Lepidoptera, Noctuidae) - a Review of Its Occurrence and Detection in 4 Pest Species. Australian Journal of Zoology 35: 1-24.

Featherstone, M. 2002. Coactivators in transcription initiation: here are your orders. Curr Opin Genet Dev 12: 149-155.

Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6-13.

Felizmenio-Quimio, M. E., N. L. Daly, and D. J. Craik. 2001. Circular proteins in plants - Solution structure of a novel macrocyclic trypsin inhibitor from Momordica cochinchinensis. Journal of Biological Chemistry 276: 22875-22882.

Ferguson, D. L., R. B. Turley, and R. H. Kloth. 1997. Identification of a delta-TIP cDNA clone and determination of related A and D genome subfamilies in Gossypium species. Plant Molecular Biology 34: 111-118.

Fessele, S., H. Maier, C. Zischek, P. Nelson, and T. Werner. 2002. Regulatory context is a crucial part of gene function. Trends Genet. 18: 60-3.

Fiete, D. J., M. C. Beranek, and J. U. Baenziger. 1998. A cysteine-rich domain of the "mannose" receptor mediates GalNAc-4-SO4 binding. Proceedings of the National Academy of Sciences of the United States of America 95: 2089-2093.

Fitt, G. 1994. Cotton Pest Management: Part 3. An Australian Perspective. Ann. Rev. Entomol.: 543-562.

Fitt, G., L. Wilson, R. Mensah, and J. Daly. 2004. Advances with Integrated Pest Management as a component of sustainable agriculture: the case of the Australian cotton industry, New directions for a diverse planet: Proceedings of the 4th International Crop Science Congress, Brisbane, Australia.

Fitt, G. P. 2004a. Implementation and impact of transgenic Bt cottons in Australia, pp. 371-381, Third World Cotton Research Conference: Cotton Production for the New Millenium. Agricultural Research Council - Institute for Industrial Crops, Cape Town, South Africa.

Fitt, G. P. 2004b. Global status and impacts of biotech cotton: Report of the Second Expert Panel on Biotechnology of Cotton, pp. 72. ICAC, Washington.

Fitt, G. P., J. C. Daly, C. L. Mares, and O. K. 1998. Changing Efficacy of Transgenic Bt Cotton - Patterns and Consequences, pp. 189-196, Pest Management - Future Challenges: Proceedings of the 6th Australasian Applied Entomology Conference. University of Qld Press, Brisbane.

Franco, O. L., D. J. Rigden, F. R. Melo, and M. F. Grossi-de-Sá. 2002. Plant {alpha}-amylase inhibitors and their interaction with insect {alpha}-amylases: Structure, function and potential for crop protection. Eur. J. Biochem. 269: 397-412.

Franks, T., D. G. He, and M. Thomas. 1998. Regeneration of transgenic *Vitis vinifera* L. Sultana plants: genotypic and phenotypic analysis. Molecular Breeding 4: 321-333.

Frutos, R., C. Rang, and F. Royer. 1999. Managing insect resistance to plants producing Bacillus thuringiensis toxins. Critical Reviews in Biotechnology 19: 227-276.

Fryxell, P. A. 1992. A revised taxonomic interpretation of GossypiumL. (Malvaceae). Rheedea 2: 108-165.

Fu, H. Y., S. Y. Kim, and W. D. Park. 1995. High-Level Tuber Expression and Sucrose Inducibility of a Potato Sus4 Sucrose Synthase Gene Require 5'-Flanking and 3'-Flanking Sequences and the Leader Intron. Plant Cell 7: 1387-1394.

Fujita, A., N. Soma, N. Goto-Yamamoto, H. Shindo, T. Kakuta, T. Koizumi, and K. Hashizume. 2005. Anthocyanidin reductase gene expression and accumulation of flavan-3-ols in grape berry. American Journal of Enology and Viticulture 56: 336-342.

Galau, G. A., H. W. Bass, and D. W. Hughes. 1988. Restriction fragment length polymorphisms in diploid and allotetraploid *Gossypium*: Assigning the late embryogenesis-abundant (Lea) alloalleles in *G. hirsutum*. Mol Gen Genet 211: 305-314.

Gatehouse, J. A. 2002. Plant resistance towards insect herbivores: a dynamic interaction. New Phytologist 156: 145-169.

Geigenberger, P. 2003. Regulation of sucrose to starch conversion in growing potato tubers. Journal of Experimental Botany 54: 457-465.

Geigenberger, P., and M. Stitt. 1993. Sucrose Synthase Catalyzes a Readily Reversible-Reaction Invivo in Developing Potato-Tubers and Other Plant-Tissues. Planta 189: 329-339.

Gill, S. S., E. A. Cowles, and P. V. Pietrantonio. 1992. The Mode of Action of Bacillus-Thuringiensis Endotoxins. Annual Review of Entomology 37: 615-636.

Giri, A. P., and M. S. Kachole. 1998. Amylase inhibitors of pigeonpea (Cajanus cajan) seeds. Phytochemistry 47: 197-202.

Goto-Yamamoto, N., G. H. Wan, K. Masaki, and S. Kobayashi. 2002. Structure and transcription of three chalcone synthase genes of grapevine (Vitis vinifera). Plant Science 162: 867-872.

Gould, F. 1994. Potential and Problems with High-Dose Strategies for Pesticidal Engineered Crops. Biocontrol Science and Technology 4: 451-461.

Gould, F. 1998. Sustainability of transgenic insecticidal cultivars: Integrating pest genetics and ecology. Annual Review of Entomology 43: 701-726.

Greenplate, J. T. 1999. Quantification of Bacillus thuringiensis insect control protein Cry1Ac over time in Bollgard cotton fruit and terminals. Journal of Economic Entomology 92: 1377-1383.

Gurr, S. J., and P. J. Rushton. 2005. Engineering plants with increased disease resistance: how are we going to express it? Trends in Biotechnology 23: 283-290.

Gutierrez-Alcala, G., L. Calo, F. Gros, J. C. Caissard, C. Gotor, and L. C. Romero. 2005. A versatile promoter for the expression of proteins in glandular and non-glandular trichomes from a variety of plants. Journal of Experimental Botany 56: 2487-2494.

Haigler, C. H., M. Ivanova-Datcheva, P. S. Hogan, V. V. Salnikov, S. Hwang, K. Martin, and D. P. Delmer. 2001. Carbon partitioning to cellulose synthesis. Plant Molecular Biology 47: 29-51.

Harmer, S. E. 2003. Characterisation of alpha-expansin genes in Gossypium hirsutum, School of Molecular and Biomedical Science (Discipline of Genetics). University of Adelaide, Adelaide.

Harmer, S. E., S. J. Orford, and J. N. Timmis. 2002. Characterisation of six alpha-expansin genes in Gossypium hirsutum (upland cotton). Molecular Genetics and Genomics 268: 1-9.

Hearn, A., and G. Fitt. 1992. Cotton Cropping Systems. In: Field Crop Ecosystems. Elsevier, London.

Hernandez, J. F., J. Gagnon, L. Chiche, T. M. Nguyen, J. P. Andrieu, A. Heitz, T. T. Hong, T. T. C. Pham, and D. L. Nguyen. 2000. Squash trypsin inhibitors from Momordica cochinchinensis exhibit an atypical macrocyclic structure. Biochemistry 39: 5722-5730.

Higo, K., Y. Ugawa, M. Iwamoto, and T. Korenaga. 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res 27: 297-300.

Hilder, V. A., and D. Boulter. 1999. Genetic engineering of crop plants for insect resistance - a critical review. Crop Protection 18: 177-191.

Hirose, T., and T. Terao. 2004. A comprehensive expression analysis of the starch synthase gene family in rice (Oryza sativa L.). Planta 220: 9-16.

Hodgson, C. P., and R. Z. Fisk. 1987. Hybridization probe size control: optimized 'oligolabelling'. Nucleic Acids Res 15: 6295.

Hofmann, C., H. Vanderbruggen, H. Hofte, J. Vanrie, S. Jansens, and H. Vanmellaert. 1988. Specificity of Bacillus-Thuringiensis Sigma-Endotoxins Is Correlated with the Presence of High-Affinity Binding-Sites in the Brush-Border Membrane of Target Insect Midguts. Proceedings of the National Academy of Sciences of the United States of America 85: 7844-7848.

Horsefall, J. 1972. genetic Vulnerability of Major Crops. National Academy of Sciences, Washington DC.

Hsu, C. Y., R. G. Creech, J. N. Jenkins, and D. P. Ma. 1999. Analysis of promoter activity of cotton lipid transfer protein gene LTP6 in transgenic tobacco plants. Plant Science 143: 63-70.

Hudspeth, R. L., and J. W. Grula. 1989. Structure and Expression of the Maize Gene Encoding the Phosphoenolpyruvate Carboxylase Isozyme Involved in C-4 Photosynthesis. Plant Molecular Biology 12: 579-589.

Husnain, T., J. Asad, S. B. Maqbool, S. K. Datta, and S. Riazuddin. 2002. Variability in expression of insecticidal Cry1Ab gene in Indica Basmati rice. Euphytica 128: 121-128.

Inoue, H., H. Nojima, and H. Okayama. 1990. High efficiency transformation of Escherichia coli with plasmids. Gene 96: 23-8.

Ishimoto, M., and M. J. Chrispeels. 1996. Protective mechanism of the Mexican bean weevil against high levels of alpha-amylase inhibitor in the common bean. Plant Physiology 111: 393-401.

Jahne, A., C. Fritzen, and G. Weissenbock. 1993. Chalcone Synthase and Flavonoid Products in Primary-Leaf Tissues of Rye and Maize. Planta 189: 39-46.

Jefferson, R. A., T. A. Kavanagh, and M. W. Bevan. 1987. Gus Fusions - Beta-Glucuronidase as a Sensitive and Versatile Gene Fusion Marker in Higher-Plants. Embo Journal 6: 3901-3907.

Jennings, C., J. West, C. Waine, D. Craik, and M. Anderson. 2001. Biosynthesis and insecticidal properties of plant cyclotides: The cyclic knotted proteins from Oldenlandia affinis. Proceedings of the National Academy of Sciences of the United States of America 98: 10614-10619.

Jennings, C. V., K. J. Rosengren, N. L. Daly, M. Plan, J. Stevens, M. J. Scanlon, C. Waine, D. G. Norman, M. A. Anderson, and D. J. Craik. 2005. Isolation, solution structure, and insecticidal activity of Kalata B2, a circular protein with a twist: Do Mobius strips exist in nature? Biochemistry 44: 851-860.

Jiang, C., R. J. Wright, K. M. El-Zik, and A. H. Paterson. 1998. Polyploid formation created unique avenues for response to selection in gossypium. Proc Natl Acad Sci U S A 95: 4419-24.

John, M. E. 1992. An efficient method for isolation of RNA and DNA from plants containing polyphenolics. Nucl Acids Res 20: 2381.

John, M. E. 1997. Cotton crop improvement through genetic engineering. Critical Reviews in Biotechnology 17: 185-208.

Johnson, K. D., H. Hofte, and M. J. Chrispeels. 1990. An Intrinsic Tonoplast Protein of Protein Storage Vacuoles in Seeds Is Structurally Related to a Bacterial Solute Transporter (Glpf). Plant Cell 2: 525-532.

Johnson, M. A., E. J. Baker, J. T. Colbert, and P. J. Green. 1998. Determinants of mRNA Stability in Plants in a Look Beyond Transcription: Mechanisms Determining mRNA Stability and Translation in Plants. American Society of Plant Physiologists.

Joshi, C. P. 1987a. Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. Nucl Acids Res 15: 9627-9640.

Joshi, C. P. 1987b. An inspection of the domain between the putative TATA box and the translation start site in 79 plant genes. Nucleic Acids Res 15: 6643-6653.

Joshi, C. P., H. Zhou, X. Huang, and V. L. Chiang. 1997. Context sequences of translation initiation codon in plants. Plant Mol Biol 35: 993-1001.

Junghans, H., K. Dalkin, and R. A. Dixon. 1993. Stress Responses in Alfalfa (Medicago-Sativa L) .15. Characterization and Expression Patterns of Members of a Subset of the Chalcone Synthase Multigene Family. Plant Molecular Biology 22: 239-253.

Kaiser, T., and A. Batschauer. 1995. Cis-Acting Elements of the Chs1 Gene from White Mustard Controlling Promoter Activity and Spatial Patterns of Expression. Plant Molecular Biology 28: 231-243.

Katterman, F. R. H., and V. I. Shattuck. 1983. An effective method of DNA isolation from the mature leaves of *Gossypium* species that contain large amounts of phenolic terpenoids and tannins. Prep Biochem 13: 347-359.

Kennedy, C. W., W. C. Smith, and J. E. Jones. 1986. Effect of Early Season Square Removal on 3 Leaf Types of Cotton. Crop Science 26: 139-145.

Kim, H. J., M. Y. Williams, and B. A. Triplett. 2002. A novel expression assay system for fibre-specific promoters in developing cotton fibres. Plant Molecular Biology Reporter 20: 7-18.

Kirschman, J. A., and J. H. Cramer. 1988. Two new tools: multi-purpose cloning vectors that carry kanamycin or spectinomycin/streptomycin resistance markers. Gene 68: 163-5.

Koes, R., W. Verweij, and F. Quattrocchio. 2005. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends in Plant Science 10: 236-242.

Kost, B., and N. H. Chua. 2002. The plant cytoskeleton: Vacuoles and cell walls make the difference. Cell 108: 9-12.

Koziel, M., N. Desai, G. Warren, N. Carozzi, R. Crenshaw, K. Launis, M. Wright, M. Meghji, G. Beland, and S. Evola. 1993. Transgenic Maize Plants for the Control of European Corn-Borer. Abstracts of Papers of the American Chemical Society 206: 120-AGRO.

Lawton, M. A., S. D. Clouse, and C. J. Lamb. 1990. Glutathione-Elicited Changes in Chromatin Structure within the Promoter of the Defense Gene Chalcone Synthase. Plant Cell Reports 8: 561-564.

Lessard, P. A., H. Kulaveerasingam, G. M. York, A. Strong, and A. J. Sinskey. 2002. Manipulating gene expression for the metabolic engineering of plants. Metabolic Engineering 4: 67-79.

Li, X. B., X. P. Fan, X. L. Wang, L. Cai, and W. C. Yang. 2005. The cotton ACTIN1 gene is functionally expressed in fibers and participates in fiber elongation. Plant Cell 17: 859-875.

Liu, L., D. C. Gitz, and J. W. McClure. 1995. Effects of Uv-B on Flavonoids, Ferulic Acid, Growth and Photosynthesis in Barley Primary Leaves. Physiologia Plantarum 93: 725-733.

Liu, M. S., F. Wang, Y. X. Dong, and Y. S. Zhang. 2005. Expression analysis of dihydroflavonol 4-reductase genes involved in anthocyanin biosynthesis in purple grains of wheat. Journal of Integrative Plant Biology 47: 1107-1114.

Lo Piero, A. R., I. Puglisi, and G. Petrone. 2006. Gene characterization, analysis of expression and in vitro synthesis of dihydroflavonol 4-reductase from Citrus sinensis (L.) Osbeck. Phytochemistry 67: 684-695.

Loake, G. J., O. Faktor, C. J. Lamb, and R. A. Dixon. 1992. Combination of H-Box Cctacc(N)7ct and G-Box (Cacgtg) Cis Elements Is Necessary for Feedforward Stimulation of a Chalcone Synthase Promoter by the Phenylpropanoid-Pathway Intermediate P-Coumaric Acid. Proceedings of the National Academy of Sciences of the United States of America 89: 9230-9234.

Ludevid, D., H. Hofte, E. Himelblau, and M. J. Chrispeels. 1992. The Expression Pattern of the Tonoplast Intrinsic Protein Gamma-Tip in Arabidopsis-Thaliana Is Correlated with Cell Enlargement. Plant Physiology 100: 1633-1639.

Maas, C., J. Laufs, S. Grant, C. Korfhage, and W. Werr. 1991. The Combination of a Novel Stimulatory Element in the 1st Exon of the Maize Shrunken-1 Gene with the Following Intron-1 Enhances Reporter Gene-Expression up to 1000-Fold. Plant Molecular Biology 16: 199-207.

Macintosh, S. C., T. B. Stone, S. R. Sims, P. L. Hunst, J. T. Greenplate, P. G. Marrone, F. J. Perlak, D. A. Fischhoff, and R. L. Fuchs. 1990. Specificity and Efficacy of Purified Bacillus-Thuringiensis Proteins against Agronomically Important Insects. Journal of Invertebrate Pathology 56: 258-266.

Maeshima, M. 1990. Development of Vacuolar Membranes During Elongation of Cells in Mung Bean Hypocotyls. Plant and Cell Physiology 31: 311-317.

Maeshima, M. 1992. Characterization of the Major Integral Protein of Vacuolar Membrane. Plant Physiology 98: 1248-1254.

Maeshima, M., I. Haranishimura, Y. Takeuchi, and M. Nishimura. 1994. Accumulation of Vacuolar H+-Pyrophosphatase and H+-Atpase During Reformation of the Central Vacuole in Germinating Pumpkin Seeds. Plant Physiology 106: 61-69.

Maizel, J. V., Jr., and R. P. Lenk. 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc Natl Acad Sci U S A 78: 7665-9.

Mang, H. G., J. H. Lee, J. A. Park, J. Pyee, H. S. Pai, and W. T. Kim. 2004. The CaPRP1 gene encoding a putative proline-rich glycoprotein is highly expressed in rapidly elongating early roots and leaves in hot pepper (Capsicum annuum L. cv. Pukang). Biochimica Et Biophysica Acta-General Subjects 1674: 103-108.

Marrone, P. G., and S. C. MacIntosh. 1993. Resistance to Bacillus thuringiensis and Resistance Management. In: Bacillus thuringiensis, An Environmental Biopesticide: Theory and Practice. John Wiley & Sons, Chichester, UK.

Mathur, J., P. Spielhofer, B. Kost, and N. H. Chua. 1999. The actin cytoskeleton is required to elaborate and maintain spatial patterning during trichome cell morphogenesis in Arabidopsis thaliana. Development 126: 5559-5568.

Matthews, G. A., and J. P. Tunstall. 1994. Insect pests of cotton. CAB International, Wallingford, Oxon.

McClintock, J. T., C. R. Schaffer, and R. D. Sjoblad. 1995. A comparative review of the mammalian toxicity of *Bacillus thuriensis*-based pesticides. Pesticide Science 45: 95-105.

Mehlo, L., D. Gahakwa, P. T. Nghia, N. T. Loc, T. Capell, J. A. Gatehouse, A. M. R. Gatehouse, and P. Christou. 2005. An alternative strategy for sustainable pest resistance in genetically enhanced crops. Proceedings of the National Academy of Sciences of the United States of America 102: 7812-7816.

Meinert, M. C., and D. P. Delmer. 1977. Changes in biochemical composition of the cell wall of the cotton fibre during development. Plant Physiol 59: 1088-1097.

Meyer, P. 2000. Transcriptional transgene silencing and chromatin components. Plant Mol Biol 43: 221-234.

Michaud, D. 1997. Avoiding protease-mediated resistance in herbivorous pests. Trends in Biotechnology 15: 4-6.

Moulherat, C., M. Tengberg, J.-F. Haquet, and B. Mille. 2002. First Evidence of Cotton at Neolithic Mehrgarh, Pakistan: Analysis of Mineralized Fibres from a Copper Bead. Journal of Archaeological Science 29: 1393-1401.

Nakamura, M., T. Tsunoda, and J. Obokata. 2002. Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. Plant Journal 29: 1-10.

Nothnagel, E. A. 1997. Proteoglycans and related components in plant cells, pp. 195-291, International Review of Cytology - a Survey of Cell Biology, Vol 174.

O'Callaghan, M., T. R. Glare, E. P. J. Burgess, and L. A. Malone. 2005. Effects of plants genetically modified for insect resistance on nontarget organisms. Annual Review of Entomology 50: 271-292.

Odell, J. T., F. Nagy, and N. H. Chua. 1985. Identification of DNA-Sequences Required for Activity of the Cauliflower Mosaic Virus-35s Promoter. Nature 313: 810-812.

Oerke, E., H.-W. Dehne, F. Schonbeck, and A. Weber. 1994. Crop Production and Crop Protection: Estimated losses in major food and cash crops. Elsevier, Amsterdam.

Pannetier, C., M. Giband, P. Couzi, V. LeTan, M. Mazier, J. Tourneur, and B. Hau. 1997. Introduction of new traits into cotton through genetic engineering: insect resistance as example. Euphytica 96: 163-166.

Pao, G. M., L. F. Wu, K. D. Johnson, H. Hofte, M. J. Chrispeels, G. Sweet, N. N. Sandal, and M. H. Saier. 1991. Evolution of the Mip Family of Integral Membrane-Transport Proteins. Molecular Microbiology 5: 33-37.

Patron, N. J., A. M. Smith, B. F. Fahy, C. M. Hylton, M. J. Naldrett, B. G. Rossnagel, and K. Denyer. 2002. The altered pattern of amylose accumulation in the endosperm of low-amylose barley cultivars is attributable to a single mutant allele of granule-bound starch synthase I with a deletion in the 5 '-non-coding region. Plant Physiology 130: 190-198.

Perlak, F. J., R. W. Deaton, T. A. Armstrong, R. L. Fuchs, S. R. Sims, J. T. Greenplate, and Fischhoff. 1990. Insect resistant cotton plants. Biotech 8: 939-943.

Perlak, F. J., M. Oppenhuizen, K. Gustafson, R. Voth, S. Sivasupramaniam, D. Heering, B. Carey, R. A. Ihrig, and J. K. Roberts. 2001. Development and commercial use of Bollgard cotton in the USA--early promises versus today's reality. Plant J 27: 489-501.

Peters, D. J., and C. P. Constabel. 2002. Molecular analysis of herbivore-induced condensed tannin synthesis: cloning and expression of dihydroflavonol reductase from trembling aspen (Populus tremuloides). Plant Journal 32: 701-712.

Pimentel, D., and M. Burgess. 1985. Effects of Single Versus Combinations of Insecticides on the Development of Resistance. Environmental Entomology 14: 582-589.

Prestridge, D. S. 1991. Signal Scan - a Computer-Program That Scans DNA-Sequences for Eukaryotic Transcriptional Elements. Computer Applications in the Biosciences 7: 203-206.

Prutz, G., A. Brink, and K. Dettner. 2004. Transgenic insect-resistant corn affects the fourth trophic level: effects of Bacillus thuringiensis-corn on the facultative hyperparasitoid Tetrastichus howardi. Naturwissenschaften 91: 451-454.

Pyke, B. A., and E. H. Brown. 1996. The Cotton Pest and Beneficial Guide. CRDC, Narrabri.

Raven, P. H., R. F. Evert, and S. E. Eichorn. 1992. Biology of Plants. Worth Publishers, NY.

Reddy, V. S., and A. S. N. Reddy. 2004. Developmental and cell-specific expression of ZWICHEL is regulated by the intron and exon sequences of its upstream protein-coding gene. Plant Molecular Biology 54: 273-293.

Reuber, S., J. F. Bornman, and G. Weissenbock. 1996. Phenylpropanoid compounds in primary leaf tissues of rye (Secale cereale). Light response of their metabolism and the possible role in UV-B protection. Physiologia Plantarum 97: 160-168.

Richard, S., G. Lapointe, R. G. Rutledge, and A. Seguin. 2000. Induction of chalcone synthase expression in white spruce by wounding and jasmonate. Plant and Cell Physiology 41: 982-987.

Roitsch, T. 1999. Source-sink regulation by sugar and stress. Current Opinion in Plant Biology 2: 198-206.

Rombauts, S., K. Florquin, M. Lescot, K. Marchal, P. Rouze, and Y. Van de Peer. 2003. Computational approaches to identify promoters and cis-regulatory elements in plant genomes. Plant Physiol 132: 1162-1176.

Rose, A. B. 2004. The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis. Plant Journal 40: 744-751.

Rose, A. B., and J. A. Beliakoff. 2000. Intron-mediated enhancement of gene expression independent of unique intron sequences and splicing. Plant Physiology 122: 535-542.

Ross, H. A., H. V. Davies, L. R. Burch, R. Viola, and D. McRae. 1994. Developmental-Changes in Carbohydrate Content and Sucrose Degrading Enzymes in Tuberising Stolons of Potato (Solanum-Tuberosum). Physiologia Plantarum 90: 748-756.

Roush, R. T. 1997. Bt-transgenic crops: just another pretty insecticide or a chance for a new start in resistance management? Pesticide Science 51: 328-334.

Roush, R. T. 1998. Two-toxin strategies for management of insecticidal transgenic crops: can pyramiding succeed where pesticide mixtures have not? Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 353: 1777-1786.

Roush, R. T., and A. M. Shelton. 1997. Assessing the odds: The emergence of resistance to bt transgenic plants. Nature Biotechnology 15: 816-817.

Ruan, Y. L., and P. S. Chourey. 1998. A fiberless seed mutation in cotton is associated with lack of fiber cell initiation in ovule epidermis and alterations in sucrose synthase expression and carbon partitioning in developing seeds. Plant Physiol 118: 399-406.

Ruan, Y. L., D. J. Llewellyn, and R. T. Furbank. 2001. The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K+ transporters and expansin. Plant Cell 13: 47-60.

Ruan, Y. L., D. J. Llewellyn, R. T. Furbank, and P. S. Chourey. 2005. The delayed initiation and slow elongation of fuzz-like short fibre cells in relation to altered patterns of sucrose synthase expression and plasmodesmata gating in a lintless mutant of cotton. Journal of Experimental Botany 56: 977-984.

Sadras, V. O. 1995. Compensatory Growth in Cotton after Loss of Reproductive-Organs. Field Crops Research 40: 1-18.

Sakuta, M. 2000. Transcriptional control of chalcone synthase by environmental stimuli. Journal of Plant Research 113: 327-333.

Salehuzzaman, S., E. Jacobsen, and R. G. F. Visser. 1993. Isolation and Characterization of a Cdna-Encoding Granule-Bound Starch Synthase in Cassava (Manihot-Esculenta Crantz) and Its Antisense Expression in Potato. Plant Molecular Biology 23: 947-962.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual.*

Sargent, T. D. 1987. Isolation of differentially expressed genes. Methods Enzymol 152: 423-432.

Saxena, D., and G. Stotzky. 2000. Insecticidal toxin from Bacillus thuringiensis is released from roots of transgenic Bt corn in vitro and in situ. Fems Microbiology Ecology 33: 35-39.

Schneiderbauer, A., H. Sandermann, and D. Ernst. 1991. Isolation of Functional Rna from Plant-Tissues Rich in Phenolic-Compounds. Analytical Biochemistry 197: 91-95.

Schnepf, H. E., and H. R. Whiteley. 1981. Cloning and Expression of the Bacillus-Thuringiensis Crystal Protein Gene in Escherichia-Coli. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 78: 2893-2897.

Schubert, A. M., C. R. Benedict, J. D. Berlin, and R. J. Kohel. 1973. Cotton fibre development-kinetics of cell elongation and secondary wall thickening. Crop Sci 13: 704-709.

Schuler, T. H., G. M. Poppy, B. R. Kerry, and I. Denholm. 1998. Insect-resistant transgenic plants. Trends in Biotechnology 16: 168-175.

Schunmann, P. H. D., D. J. Llewellyn, B. Surin, P. Boevink, F. R. C. De, and P. M. Waterhouse. 2003. A suite of novel promoters and terminators for plant biotechnology. Functional Plant Biology 30: 443-452.

Sequeira, R. 2001. Inter-seasonal population dynamics and cultural management of Helicoverpa spp. in a Central Queensland cropping system. Australian Journal of Experimental Agriculture 41: 249 - 259.

Sharma, H. C., K. K. Sharma, and J. H. Crouch. 2004. Genetic transformation of crops for insect resistance: Potential and limitations. Critical Reviews in Plant Sciences 23: 47-72.

Sharma, H. C., K. K. Sharma, N. Seetharama, and R. Ortiz. 2000. Prospects for using transgenic resistance to insects in crop improvement. Electronic Journal of Biotechnology 3.

Shaw, J. R., R. J. Ferl, J. Baier, D. Stclair, C. Carson, D. R. McCarty, and L. C. Hannah. 1994. Structural Features of the Maize Sus1 Gene and Protein. Plant Physiology 106: 1659-1665.

Shelton, A. M., J. Z. Zhao, and R. T. Roush. 2002. Economic, ecological, food safety, and social consequences of the deployment of Bt transgenic plants. Annual Review of Entomology 47: 845-881.

Sheng, J., R. D'Ovidio, and M. C. Mehdy. 1991. Negative and positive regulation of a novel proline-rich mRNA by fungal elicitor and wounding. Plant J 1: 345-354.

Shimada, N., R. Sasaki, S. Sato, T. Kaneko, S. Tabata, T. Aoki, and S. Ayabe. 2005. A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the Lotus japonicus genome. Journal of Experimental Botany 56: 2573-2585.

Showalter, A. M. 2001. Arabinogalactan-proteins: structure, expression and function. Cellular and Molecular Life Sciences 58: 1399-1417.

Simpson, G. G., and W. Filipowicz. 1996. Splicing of precursors to mRNA in higher plants: Mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery. Plant Molecular Biology 32: 1-41.

Sims, S. R., and S. A. Berberich. 1996. Bacillus thuringiensis CryIA protein levels in raw and processed seed of transgenic cotton: Determination using insect bioassay and ELISA. Journal of Economic Entomology 89: 247-251.

Sivamani, E., and R. Qu. 2006. Expression enhancement of a rice polyubiquitin gene promoter. Plant Molecular Biology 60: 225-239.

Smale, S. T., and D. Baltimore. 1989. The Initiator as a Transcription Control Element. Cell 57: 103-113.

Smale, S. T., A. Jain, J. Kaufmann, K. H. Emami, K. Lo, and I. P. Garraway. 1998. The initiator element: A paradigm for core promoter heterogeneity within metazoan protein-coding genes. Cold Spring Harbor Symposia on Quantitative Biology 63: 21-31.

Smart, L. B., F. Vojdani, M. Maeshima, and T. A. Wilkins. 1998. Genes involved in osmoregulation during turgor-driven cell expansion of developing cotton fibers are differentially regulated. Plant Physiol 116: 1539-49.

Smart, L. B., W. A. Moskal, K. D. Cameron, and A. B. Bennett. 2001. MIP genes are down-regulated under drought stress in Nicotiana glauca. Plant and Cell Physiology 42: 686-693.

Smith, A. M., K. Denyer, and C. Martin. 1997. The synthesis of the starch granule. Annual Review of Plant Physiology and Plant Molecular Biology 48: 65-87.

Steimer, A., H. Schob, and U. Grossniklaus. 2004. Epigenetic control of plant development: new layers of complexity. Current Opinion in Plant Biology 7: 11-19.

Stephens, S. G., and M. E. Moseley. 1974. Early domesticated cottons from archaeological sites in central coastal Peru. Am. Antiq. 39: 109–122.

Sunilkumar, G., L. Mohr, F. E. Lopata, C. Emani, and K. S. Rathore. 2002a. Developmental and tissue-specific expression of CaMV 35S promoter in cotton as revealed by GFP. Plant Molecular Biology 50: 463-474.

Sunilkumar, G., J. P. Connell, C. W. Smith, A. S. Reddy, and K. S. Rathore. 2002b. Cotton alpha-globulin promoter: Isolation and functional characterization in transgenic cotton, Arabidopsis, and tobacco. Transgenic Research 11: 347-359.

Tabashnik, B. E., F. Gould, and Y. Carriere. 2004. Delaying evolution of insect resistance to transgenic crops by decreasing dominance and heritability. Journal of Evolutionary Biology 17: 904-912.

Takeuchi, A., S. Matsumoto, and M. Hayatsu. 1994. Chalcone Synthase from Camellia-Sinensis - Isolation of the Cdnas and the Organ-Specific and Sugar-Responsive Expression of the Genes. Plant and Cell Physiology 35: 1011-1018.

Takos, A. M., B. E. Ubi, S. P. Robinson, and A. R. Walker. 2006. Condensed tannin biosynthesis genes are regulated separately from other flavonoid biosynthesis genes in apple fruit skin. Plant Science 170: 487-499.

Tanner, G. J., K. T. Francki, S. Abrahams, J. M. Watson, P. J. Larkin, and A. R. Ashton. 2003. Proanthocyanidin biosynthesis in plants - Purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. Journal of Biological Chemistry 278: 31647-31656.

Tenorio, G., A. Orea, J. M. Romero, and A. Merida. 2003. Oscillation of mRNA level and activity of granule-bound starch synthase I in Arabidopsis leaves during the day/night cycle. Plant Molecular Biology 51: 949-958.

Tetlow, I. J., M. K. Morell, and M. J. Emes. 2004. Recent developments in understanding the regulation of starch metabolism in higher plants. Journal of Experimental Botany 55: 2131-2145.

Thomas, J. C., D. G. Adams, V. D. Keppenne, C. C. Wasmann, J. K. Brown, M. R. Kanost, and H. J. Bohnert. 1995. Protease inhibitors of Manduca sexta expressed in transgenic cotton. Plant Cell Reports 14: 758-762.

Thompson, D. B. 2000. On the non-random nature of amylopectin branching. Carbohydrate Polymers 43: 223-239.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-80.

Tian, G.-W., A. Mohanty, S. N. Chary, S. Li, B. Paap, G. Drakakaki, C. D. Kopec, J. Li, D. Ehrhardt, D. Jackson, S. Y. Rhee, N. V. Raikhel, and V. Citovsky. 2004. High-throughput fluorescent tagging of full-length Arabidopsis gene products in planta. Plant Physiol 135: 1-14.

Tian, L., M. P. Fong, J. Y. J. Wang, N. E. Wei, H. M. Jiang, R. W. Doerge, and Z. J. Chen. 2005. Reversible histone acetylation and deacetylation mediate genome-wide, promoter-dependent and locus-specific changes in gene expression during plant development. Genetics 169: 337-345.

Tompa, M., N. Li, T. L. Bailey, G. M. Church, B. De Moor, E. Eskin, A. V. Favorov, M. C. Frith, Y. Fu, W. J. Kent, V. J. Makeev, A. A. Mironov, W. S. Noble, G. Pavesi, G. Pesole, M. Regnier, N. Simonis, S. Sinha, G. Thijs, J. van Helden, M. Vandenbogaert, Z. Weng, C. Workman, C. Ye, and Z. Zhu. 2005. Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol 23: 137-44.

Tyagi, A. K. 2001. Plant genes and their expression. Current Science 80: 161-169.

Vain, P., N. Keen, J. Murillo, C. Rathus, C. Nemes, and J. J. Finer. 1993. Development of the Particle Inflow Gun. Plant Cell, Tissue and Organ Culture 33: 237-246.

Vanderleij, F. R., R. G. F. Visser, A. S. Ponstein, E. Jacobsen, and W. J. Feenstra. 1991. Sequence of the Structural Gene for Granule-Bound Starch Synthase of Potato (Solanum-Tuberosum L) and Evidence for a Single Point Deletion in the Amf Allele. Molecular & General Genetics 228: 240-248.

Vojtech, E., M. Meissle, and G. M. Poppy. 2005. Effects of Bt maize on the herbivore Spodoptera littoralis (Lepidoptera : Noctuidae) and the parasitoid Cotesta marginiventris (Hymenoptera : Braconidae). Transgenic Research 14: 133-144.

Wagh, P. V., and O. P. Bahl. 1981. Sugar Residues on Proteins. Crc Critical Reviews in Biochemistry 10: 307-377.

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

Wendel, J. F. 1989. New World tetraploid cottons contain Old World cytoplasm. Proc Natl Acad Sci USA 86: 4132-4136.

Wilkins, T. A., and J. A. Jernstedt. 1999. Molecular genetics of developing cotton fibers, pp. 231-267. *In* A. M. Basra [ed.], Cotton Fibers. Hawthorne Press, New York.

Wilkins, T. A., K. Rajasekaran, and D. M. Anderson. 2000. Cotton Biotechnology. Critical Reviews in Plant Sciences 19: 511-550.

Willmitzer, L. 1999. Plant biotechnology: output traits - the second generation of plant biotechnology products is gaining momentum - Overview. Current Opinion in Biotechnology 10: 161-162.

Wilson, L. J. 1993. Spider mites (Acari: Tetranychidae) affect yield and fiber quality of cotton. J. Econ. Entomol 86: 566-585.

Wilson, L. J. 1994. Resistance of Okra-Leaf Cotton Genotypes to 2-Spotted Spider-Mites (Acari, Tetranychidae). Journal of Economic Entomology 87: 1726-1735.

Wilson, L. J., V. O. Sadras, S. C. Heimoana, and D. Gibb. 2003. How to succeed by doing nothing: Cotton compensation after simulated early season pest damage. Crop Science 43: 2125-2134.

Wu, K. M., and Y. Y. Guo. 2004. Changes in susceptibility to conventional insecticides of a Cry1Ac-selected population of Helicoverpa armigera (Hubner) (Lepidoptera : Noctuidae). Pest Management Science 60: 680-684.

Wu, Y. R., A. C. Machado, R. G. White, D. J. Llewellyn, and E. S. Dennis. 2006. Expression profiling identifies genes expressed early during lint fibre initiation in cotton. Plant and Cell Physiology 47: 107-127.

Xie, D. Y., S. B. Sharma, N. L. Paiva, D. Ferreira, and R. A. Dixon. 2003. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. Science 299: 396-399.

Yamagata, H., K. Yonesu, A. Hirata, and Y. Aizono. 2002. TGTCACA motif is a novel cis-regulatory enhancer element involved in fruit-specific expression of the cucumisin gene. Journal of Biological Chemistry 277: 11582-11590.

Yang, J., J. X. Huang, H. Y. Gu, Y. Zhong, and Z. H. Yang. 2002. Duplication and adaptive evolution of the chalcone synthase genes of dendranthema (Asteraceae). Molecular Biology and Evolution 19: 1752-1759.

Yu, W. P., A. Y. Wang, R. H. Juang, H. Y. Sung, and J. C. Su. 1992. Isolation and Sequences of Rice Sucrose Synthase Cdna and Genomic DNA. Plant Molecular Biology 18: 139-142.

Zhao, J. Z., J. Cao, Y. X. Li, H. L. Collins, R. T. Roush, E. D. Earle, and A. M. Shelton. 2003. Transgenic plants expressing two Bacillus thuringiensis toxins delay insect resistance evolution. Nature Biotechnology 21: 1493-1497.

Zwahlen, C., A. Hilbeck, R. Howald, and W. Nentwig. 2003. Effects of transgenic Bt corn litter on the earthworm Lumbricus terrestris. (vol 12, pg 1077, 2003). Molecular Ecology 12: 2279-2279.