Identification and analysis of genes involved in cotton fibre initiation

This thesis is presented for the degree of Doctor of Philosophy (PhD) in the Faculty of Sciences at the University of Adelaide

John A Humphries

Discipline of Genetics, School of Molecular and Biomedical Science, University of Adelaide

September 2006

Table of Contents

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.

Signed

John Humphries

September, 2006

Acknowledgements

I would like to thank Jeremy Timmis for his continuing encouragement during my time in the lab. When my experiments were looking a bit grim, your optimistic attitude and positive frame of mind (at least outwardly!) along with a sense of humour always made things seem better. Many thanks must also go to Sharon Orford, who did the work leading into this project, and provided great feedback and support – plus always good for a footy chat, even if you did go for the wrong team! Thanks heaps Shazz.

I am grateful to the Cotton Research and Development Corporation (CRDC) for their generous funding of this project, and for providing support for conference travel. Thanks also go to Mandy Walker at the CSIRO for her efforts in our collaboration and the use of growth facilities, and to Justine Chambers for her assistance with Arabidopsis growth and transformation.

Thanks especially need to go to current and past lab members who made the Timmis lab such an enjoyable place to work - in particular to my fellow PhD students Sas, Damien, Sven, and now Anna. I am grateful to the post-docs that have been in the lab, namely Chunyuan and Eun-Lee, from whom I have learnt much and enjoyed their company. The JNT honours students during my time have also been great, and have added something different to the lab each year! I have been privileged to work with and get to know such a fantastic group of people.

I am very appreciative of the people around the department who have made my PhD experience a memorable one, and have always made for a great night out - Tash, Jo, Simon, Ben, Jai, Catherine, Clare, Mandy and Ainslie to name a few. Thanks also to my non-lab friends, in particular Luke, Adam, Barbi and the K-man, for sharing many good times in the past few years. I have also enjoyed my experiences with the AUTC touch footy team and of course the mighty Van Houten All-stars!

The continued encouragement of my family has been great, and their support of all my decisions has been much appreciated - I am very lucky. Finally, thanks to Marisa for putting up with me during this time, for all your help and support, and for single-handedly ensuring not a single double space exists in this thesis!

It's time to go on to new things now, it will be strange waking up and not heading straight for the JNT lab, as has been my habit for the last few years, but I look forward to the challenges ahead.

Abstract

Cotton fibres are single-celled hairs, arising from the epidermal surface of the cotton ovule. One factor in determining the length of the mature cotton fibre is the timing of fibre initiation, which is therefore a crucial step in obtaining commercial cotton fibres. To achieve a greater understanding of the regulation of cotton fibre differentiation, more fundamental information is needed on the signals and mechanisms associated with fibre initiation. The extensive genetic knowledge of *Arabidopsis* leaf trichomes could aid in the elucidation of the genetic mechanisms controlling cotton fibre differentiation. Trichomes are small hairs on the plant surface, originating from single epidermal cells in a developmental process that appears very similar to that of cotton fibres. *Arabidopsis* trichome development has been extensively investigated, and several genes that control the process have been characterised. One gene essential for trichome initiation is *TRANSPARENT TESTA GLABRA1* (*TTG1*), and loss-of-function mutations in *TTG1* result in an almost complete absence of leaf trichomes. *TTG1* plays additional roles in numerous pathways in *Arabidopsis*, including root hair initiation, anthocyanin production and seed coat mucilage production. In order to isolate genes required for fibre initiation in cotton, functional homologues of *Arabidopsis TTG1* in cotton have been sought.

Four putative homologues of *Arabidopsis TTG1* have previously been isolated in this laboratory by RT-PCR of mRNA prepared from cotton fibres, and are termed GhTTG1-4. Sequence comparisons between the four cotton deduced proteins and *Arabidopsis* TTG1 showed that they form two groups, with GhTTG1 and GhTTG3 being closely related to each other (87% identical and 93% similar) and to TTG1 (79% and 80% amino acid identity respectively). GhTTG2 and GhTTG4 formed the second group, with 95% amino acid identity to each other and lower (approximately 62%) identity to TTG1. An analysis of the genomic origins of the *GhTTG* genes demonstrated that each is derived from the same ancestral diploid genome.

Cross-species complementation experiments were performed to test for functional homology of these cotton *TTG1*-like genes to *AtTTG1*, by introducing the cotton genes into *Arabidopis ttg1-1* mutants via *Agrobacterium*-mediated transformation. This experiment

showed that two of the four genes, *GhTTG1* and *GhTTG3* were able to restore trichome initiation in the *Arabidopsis* mutant plants, and a further investigation of *GhTTG3* transgenic plants demonstrated complementation of the full range of *ttg1* mutant phenotypes.

An analysis of the temporal and spatial expression of the *GhTTG* genes in cotton is also described. It was shown that each of the genes is expressed ubiquitously throughout the cotton plant, in common with many plant WD-repeat genes. A closer examination of transcript abundance in the developing cotton ovule utilising *in situ* hybridisation revealed predominant expression of *GhTTG1/GhTTG3* in the epidermal cells destined to become cotton fibres.

A yeast two-hybrid assay was utilised to identify transcription factors that may interact with GhTTG3 during fibre development. This experiment identified three cotton fibre cDNAs encoding putative interacting proteins, including one with a similar secondary structure to several TTG1-interacting proteins in *Arabidopsis*, raising the possibility of similar regulatory complexes controlling trichome initiation in *Arabidopsis* and cotton.

Chapter 1: Introduction

1.1 Cotton crop and industry

Cotton is the most widely used textile fibre in the world today, accounting for over 40% of total world fibre utilisation (Australian Agribusiness Group Review, June 2005). It was first grown in Australia over 100 years ago, but it has been only in the last 25 years or so that the industry has become of great significance to the country's economy. The forecast yield for the Australian cotton crop is 2.5 million bales in 2005/2006, with an estimated value of $1.16 billion (Australian Cotton Industry Council, February 2006). This value includes not only revenue gained from fibre production, but also value obtained from cotton's secondary products (such as seeds) sold for generation of oil and feedstock. The cotton industry therefore forms an important component of Australia's economy.

1.2 Cotton taxonomy

Cotton falls into the genus *Gossypium*, of the family Malvaceae. *Gossypium* is composed of approximately 45 diploid (2n=26) and five allotetraploid (2n=52) species (Fryxell, 1979). Diploid cottons are classified into eight genome groups (A-G, and K) based on cytological features and chromosome size (Endrizzi *et al.*, 1985). Hybridisation between the A and D genomes approximately 1.5 million years ago, followed by chromosome doubling, created the AD allotetraploid lineage, which includes *G. hirsutum* and *G. barbadense*. The diploid species *G. herbaceum* and *G. raimondii* are considered the best representatives of the ancestral A and D genome donors respectively (Endrizzi *et al.*, 1985). The tetraploid species each have approximately the same genome size, of around 2.2-2.9 Gb (Wendel *et al.*, 2002). Cultivated cotton is comprised of four main species, two of which are AA diploids (*Gossypium arboreum* and *Gossypium herbaceum*), and two of which are AADD allotetraploids (*Gossypium hirsutum* (Upland cotton) and *Gossypium barbadense*- also referred to as Egyptian or Sea Island cotton) (Endrizzi *et al.*, 1985). The allotetraploids account for the vast majority of cultivated cotton, due to their long fibre length and high yield. *G. hirsutum* is the most commonly cultivated, accounting for about 90% of world cotton production (Basra and Malik, 1984; Kohel, 1999). The fibres of *G. hirsutum* typically range in length from 2.2 to 3.0 cm (Basra and Malik, 1984) and, whilst

the fibres of *G. barbadense* are generally longer (over 3.5 cm), overall yield is lower in this latter species, hence the preference for *G. hirsutum*.

1.3 Growth of the cotton plant

Approximately 45 days pass from the time the cotton seed germinates before the first flower buds form, from which white-petal flowers eventually develop. The flowers typically open at night or early morning, and close approximately 24 hours later following self-pollination (Stewart, 1975). About two days after flowering, the petals are shed and the ovaries develop into a boll structure, enclosing the ovules. The cotton ovules inside this boll develop for a period of 40-60 days after flowering (or days post anthesis - DPA). The cotton fibres grow from the outer layer of the ovules, beginning from approximately the time of flowering. The boll enlarges greatly between 0 DPA and 20 DPA as the fibres inside elongate. At approximately 60 DPA, the boll splits open, exposing the fibres and facilitating harvest (Figure 1.1).

In addition to their commercial importance, the cotton fibre provides a synchronous single-cell system to study cell differentiation and elongation. Section 1.4 provides a more detailed description of the process of fibre development.

1.4 Cotton fibre development

Cotton fibres are specialised epidermal hairs derived from the ovule. Fibre development is comprised of four main phases: initiation, elongation, secondary wall synthesis and maturation, which are characterised by specific cellular events (Wilkins and Jernstedt, 1999).

1.4.1 Fibre initiation

Fibre cells initiate as single cells from the epidermal layer of the cotton ovule. Approximately 25% of epidermal cells differentiate into fibres (Beasley, 1975). Epidermal cells receive as yet unknown molecular directions to develop into fibre cells at approximately two days pre-anthesis. During the time period from two days pre-anthesis to the day of anthesis, the fibre cells are latent, awaiting stimulation from plant growth factors such as auxins and gibberellins to promote differentiation into fibres (Beasley and Ting, 1973). Beazley and Ting (1974) demonstrated that auxin (indole acetic acid, IAA) and gibberellin (GA_3) promotes ovule growth and fibre initiation in ovules cultured *in vitro*.

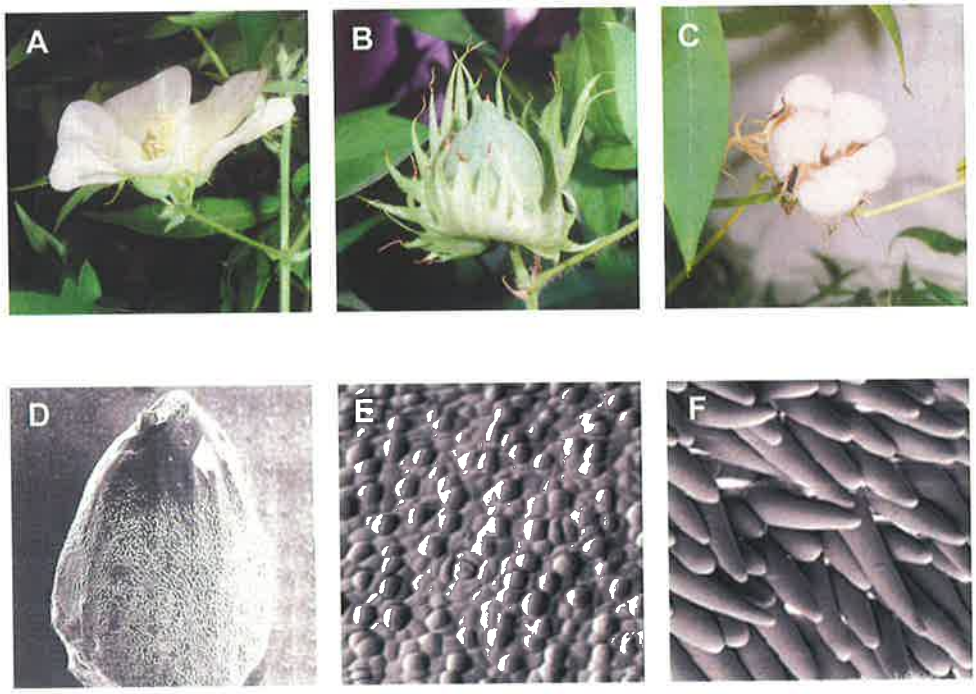


Figure 1.1: The cotton flower and fibres

The upper row **(A-C)** shows the development of the cotton flower and boll, while the lower row **(D-F)** displays SEM images of the early stages of fibre initiation. **(A)** Cotton flower on the day of anthesis. **(B)** Cotton boll at 20 DPA. **(C)** Mature fibres exposed after boll dehiscence at 50 DPA. **(D)** Ovule at the day of anthesis, with fibres initiating at the chalazal end. The arrow points to the micropylar end of the ovule, where fibres are still to initiate. Image D from Stewart, (1975). **(E)** Higher magnification view of the ovule epidermis at the day of anthesis, showing fibre initials enlarging and **(F)** at 2 DPA, where fibre cells have significantly elongated. Images E and F from Wu *et al.* (2006).

Fibre initials first begin to form at the crest of the funiculus, near the chalazal end of the ovule, at approximately 16 hours pre-anthesis (Figure 1.1D). The lateral circumference then exhibits the emergence of fibre initials, and it takes 1-4 days for fibre initiation to spread in a wave to the micropylar end of the ovule (Stewart, 1975). This wave of initiating fibres is thought to be caused by diffusion of phytohormones along the ovule epidermis. Initiating fibre cells begin to expand spherically at the day of anthesis, and the nucleus also begins to enlarge. Differentiating fibre cells have a dense cytoplasm due to both the release of phenolic substances from the vacuole and to an increased number of ribosomes present in elongating fibres (Basra and Malik, 1984). On the day of anthesis, fibre cells exhibit an increased number of Golgi complexes, likely to be involved in biosynthesis of cell wall polysaccharides (Berlin, 1986). The time-point at which initiation takes place influences the eventual length of the fibre cell (Basra and Malik, 1984). Only the longer "lint" hairs, which initiate on (or closely after) the day of anthesis are of commercial value. Several days later, another class of fibre cells initiates, which are termed fuzz. Fuzz fibres generally grow only to 1.5 - 3 millimetres, remain attached to the ovule during the ginning process, and have little commercial value (Beasley, 1973). Turley and Kloth (2002) identified several alleles that confer a fuzzless phenotype, indicating that lint and fuzz fibre types may be under separate genetic control. It is estimated that each wild-type cotton ovule produces between 13,000 and 21,000 lint fibres (Basra and Malik, 1984).

1.4.2 Fibre elongation

The elongation stage follows fibre initiation, where the cotton fibres extend by diffuse growth. After 2-3 days, the fibre cell tips become tapered and the fibres start to grow in spirals (Beasley, 1975). Elongation occurs for 20-30 days until fibres reach their final length. Fibre cells are among the longest plant cells to be studied, typically reaching lengths 1000 to 3000 times their width (Basra and Malik, 1984). During elongation, a thin primary cell wall is deposited which is 0.2 to 0.4 μm thick and has a very low cellulose concentration compared to the fibre's secondary cell wall (see Section 1.4.3 below).

1.4.3 Secondary wall synthesis

The elongation phase overlaps with secondary cell wall formation by several days, with the transition between these two phases occurring from 15 to 22 DPA (Basra and Saha, 1999).

The principal event of secondary cell wall formation is the production of cellulose. During this stage, β-1,4-glucan chains are synthesised. These chains form the cellulose microfibrils of the secondary cell wall (Delmer, 1999). The microfibrils are arranged in a helical fashion around the fibre. The secondary wall is deposited inside the primary wall, and is much thicker and stronger, defining the textile properties of the fibre. Layers of cellulose continue to be deposited until the wall is 3-4 μm thick. Rapid cellulose synthesis persists until maturation, when the fibre is almost completely composed of cellulose (Basra and Malik, 1984).

1.4.4 Maturation

Fibre maturation occurs at 50-60 DPA, when the ovary splits open, leaving seeds and mature fibres exposed to dehydration. The lint, which has twisted in such a manner that fine, yet strong threads can be spun from them, can be detached from the seeds (Basra and Malik, 1984). The mechanisms that regulate the termination of fibre development are yet to be determined, but it has been suggested that programmed cell death (PCD) might be involved, although none of the markers consistent with PCD have been identified in the cotton fibre (Lam, 2004).

1.5 Genes involved in cotton fibre development

The distinct phases of cotton fibre development are regulated by the expression of an array of genes. Significant diversity exists in the type of genes expressed throughout cotton fibre development and numerous full-length genes expressed in the fibre have been characterised. Some of these genes are expressed specifically during initiation and elongation phases, whilst others are expressed predominantly during secondary wall synthesis, and some genes remain active throughout fibre development. Complete analysis of many of these genes has been slowed by the lack of genomic sequence information, and the time and resources required for the generation of transgenic cotton. Several genes have been isolated and shown to play a functional role in fibre elongation and structural maturation, but those which regulate fibre initiation are largely unknown. A selection of the genes involved in conferring the structural properties of the fibre is discussed below. Many of these genes have been the focus of study due to their predominant expression in the fibre.

1.5.1 Genes involved in elongation and secondary cell wall growth

An analysis of elongation phase genes in cotton fibre by Ji *et al.* (2003) utilised subtractive PCR and cDNA macro-array to compare expression between wild-type and a fibreless mutant. Several cell wall proteins and enzymes were identified, including a putative vacuolar (H^+)-ATPase catalytic subunit, a kinesin-like calmodulin binding protein, some arabinogalactan proteins and enzymes involved in long chain fatty acid biosynthesis. Arpat *et al.* (2004) analysed expressed sequence tags (ESTs) from elongating cotton fibres and found that, at the time of the developmental switch from cell elongation phase to secondary cell wall syntheses, more than two thousand expansion-associated fibre genes are significantly down-regulated. In contrast, relatively few genes (81) were significantly up-regulated during secondary cell wall synthesis and these were involved in cell wall biogenesis and energy/carbohydrate metabolism. The dynamic changes in gene expression between the two phases indicate that transcription is co-ordinately regulated to facilitate each stage in the elongation process.

Genes involved in the modification of cell wall properties (cell expansion and cell wall loosening), such as 1-4-β glucanase and expansin genes, have been shown to be expressed specifically in fibre cells and allow for cell expansion and cell wall loosening (Shimizu *et al.*, 1997; Harmer *et al.*, 2002). In developing fibres, the dramatic increase in cellulose synthesis at the onset of secondary wall synthesis correlates with increased transcription of two cellulose synthase genes (*GhCesA1* and *GhCesA2*), which likely play a role in the synthesis and deposition of cellulose (Pear *et al.*, 1996). Other genes believed to play a role in cellulose synthesis are the chitinase-like genes, *GhCTL1* and *GhCTL2* (Zhang *et al.*, 2004). Several lipid transfer proteins (LTP) genes have also been shown to be expressed in cotton fibres, with a potential role in cutin deposition (Ma *et al.*, 1995; Orford and Timmis, 2000). Other genes identified as being involved in fibre development include those that encode the cytoskeleton-related α- and β-tubulins (Whittaker and Triplett, 1999; Li *et al.*, 2002b), a germin-like protein abundantly expressed in elongating fibres (Kim and Triplett, 2004), and H6, a proline-rich protein expressed predominantly during secondary wall synthesis (John and Keller, 1995).

1.5.2 Genes involved in initiation of cotton fibres

The vast majority of fibre-specific genes with functions assigned in fibre development control the structural properties of the fibre, and are generally involved in cell wall biogenesis, cytoskeleton production and energy/carbohydrate metabolism. Very few genes identified thus far have been associated with fibre initiation. One gene that may be involved in the early stages of fibre development is the *sucrose synthase* (*Sus*) gene. *Sus* participates in the conversion of sucrose from the ovule epidermis into UDP-glucose and fructose, driving cellulose synthesis and maintaining turgor in the expanding fibre cell (Ruan *et al.*, 2001). Ruan *et al.*, (2003) demonstrated that antisense-induced suppression of *Sus* inhibited fibre initiation and elongation in transgenic cotton. However, in the transgenic seeds with suppressed *Sus*, the number of protrusions (although shrunken and collapsed) from the seed coat at 0 DPA is similar to wild-type, suggesting that fibre initiation is occurring normally and the lack of *Sus* expression disrupts fibre development at a later stage. High level suppression of *Sus* activity has a more general affect on cotton seed development, leading to stunted and inviable seeds.

The most extensive analysis of fibre initiation genes thus far was performed by Wu *et al.* (2006). In this study, cDNA micro-array technology was used to compare the expression profiles of 0 DPA ovules from six fibreless mutants with wild-type ovules. The genes found to be transcriptionally down-regulated in some or all of the fibreless mutant ovules were further analysed with regard to their expression in the wild-type cotton ovule. The genes that were expressed exclusively in the outer integument of the ovule, and shown to plateau in expression after 0 DPA, were considered the primary candidates in the control of fibre initiation. These included MYB-type genes (termed GhMyb25 and GhMyb25-like), a homeodomain gene (GhHD1) and a cyclin D–related gene (GhCycD3;1). Any role these genes may have in fibre development is currently being verified by a series of transgenic cotton experiments (Wu *et al.*, 2006).

A recent study by Yang and co-workers (2006) identified numerous fibre-specific phytohormonal regulators and transcription factors enriched in *G. hirsutum* (TM-1) immature ovules (-3 to +3 DPA) compared to ovules of a seedless mutant *NIN1*, suggesting important roles for these genes in the early stages of fibre development. The putative transcription factors include MYB and homeodomain proteins, as was observed in the investigation by Wu *et al.* (2006).

An understanding of the genes involved in cotton fibre development is essential for the formulation of strategies for fibre modification, which has been a major goal in cotton biotechnology for many years.

1.6 Cotton Biotechnology

The advent of successful transgenesis in cotton and other crop plants prompted the development of agricultural biotechnology in the early 1990s (Wilkins *et al.*, 2000). Traits that may enhance agronomic performance of the cotton plant in the field include resistance to temperature and drought (abiotic stresses) and resistance to pathogens (biotic stresses).

1.6.1 Plant breeding and genetic engineering

Significant improvements in fibre production have been brought about over many years by plant breeding and selection techniques. Breeding programs are time consuming and costly and the need for hybridisation means that breeding must be restricted to compatible species capable of producing viable progeny. Thus, future improvements in fibre qualities are likely to depend largely on genetic engineering.

A potential, though often overestimated, advantage that genetic engineering possesses over conventional breeding is the reduced time required for the generation of new varieties. Genetic engineering is not confined to genetically compatible species, and hence, in combination with plant breeding methods, it allows the introduction of beneficial traits in a reduced time period. Currently, recombinant DNA technology has been used to produce a number of cotton varieties with agronomically valuable traits, including increased fibre strength, insect resistance and stress tolerance (reviewed in Wilkins *et al.*, 2000; Li *et al.*, 2004; Yan *et al.*, 2004). The widespread acceptance of transgenic cotton varieties in Australia has led to a significant reduction in production costs due to the declining use of pesticides and herbicides.

1.6.2 Towards fibre improvement in genetically modified cotton

Part of the genetic approach to improving fibre quality and yield involves identifying quantitative trait loci (QTL) associated with cotton fibre properties. A limitation placed on QTL mapping and associated plant breeding efforts is the paucity of sequence for the large cotton genome. However, a number of QTLs related to fibre characteristics have been

identified which may be utilised in selecting lines of higher fibre quality during the breeding process (Paterson *et al.*, 2003; Lin *et al.*, 2005).

The formation of the International Cotton Genome Initiative aims to develop a fully integrated physical and genetic map of the cotton genome. Rong *et al.* (2004) developed the most comprehensive map of tetraploid and diploid cottons, consisting of sequence tagged sites at 3347 loci, providing a basis for improved QTL and sequence analysis. There have also been rapid advances in gene discovery, with the release of over 45,000 ESTs from fibre, among a total of approximately 210,000 ESTs from the entire cotton genome (Arpat *et al.*, 2004; Haigler *et al.*, 2005; Udall *et al.*, 2006; Yang *et al.*, 2006), creating a vast resource to underpin fibre improvement.

1.6.3 Transformation of cotton

The initial aims of recombinant DNA technology in cotton were to improve herbicide and insect resistance. High costs are incurred as a result of cotton pest attack and the application of chemical insecticides, and thus the development of genetically engineered cotton for pest resistance has been a priority. The most widely-cultivated genetically modified (GM) cotton crops is the transgenic line expressing the endotoxin produced by the bacterium *Bacillus thuringiensis* (*Bt*), which possesses insecticidal activity against major pests (Berliner, 1915; Halcomb *et al.*, 1996). Cotton represents the first major *Bt*-resistant crop genetically engineered for commercial production in conventional agriculture. Crossing of *Bt* transgenic cotton with Roundup ready cotton has produced "stacked" varieties, containing *Bt* genes coupled with herbicide resistance genes. The uptake of these dual lines increased the proportion of transgenic cotton to 28% worldwide in 2004 (Lawrence, 2005) and to 70% of the Australian crop (Cotton Australia, 2005). Other genes have been investigated and utilised to varying extents to combat pests in cotton, including various enzyme inhibitors (Giri and Kachole, 1998; Sharma *et al.*, 2000) and cyclotides (Jennings *et al.*, 2001).

The generation of transgenic cotton can be achieved by several methods. Currently the most widely used method is somatic embryogenesis, involving *Agrobacterium*-mediated transformation of cotton tissue, leading to the eventual production of transgenic plants within approximately 12 months. The fact that many of the elite cotton varieties cannot readily be transformed means that useful traits are normally introduced first into varieties

with a higher embryonic potential, which are then backcrossed repeatedly to an elite cultivar. This process increases the overall time required to generate commercial cultivars with the desired trait.

Another major transformation method used in cotton transgenesis is microprojectile bombardment, which allows direct transformation of specific tissues, such as shoot meristems (Wilkins *et al.*, 2000). This technique involves delivery of gold or tungsten particles coated with DNA into the target plant. It too has been used to introduce marker genes (e.g. GUS) and herbicide tolerance genes, but is not as commonly used in cotton as *Agrobacterium*-mediated transformation.

Strategies to produce the desired pattern of expression, as well as improved transformation methods, are major goals for cotton biotechnology. The lengthy process of transgenesis, combined with the complications of studying a tetraploid genome, has led many researchers to seek a model system that may aid in elucidating the genetic mechanisms controlling fibre initiation (Basra, 1999). The single-celled leaf hairs (trichomes) of *Arabidopsis* have been identified as a potential model system in which to study cotton fibre differentiation.

1.7 Arabidopsis trichomes

Arabidopsis thaliana has become the major model species for plant research. It is a small annual weed of the Brassicaceae family, with ecotypes Columbia and Landsberg *erecta* most commonly used for genetic studies. *Arabidopsis* is used in a wide range of studies due to its short life cycle (six weeks from seed germination to maturation), prolific seed production, easy cultivation in restricted spaces, and small (125 Mb) genome, of which the entire sequence has been determined (*Arabidopsis* Genome Initiative, 2000).

Trichomes are hair-like structures that are found on the epidermal surfaces of almost all major groups of plants, and the branched trichomes of *Arabidopsis* have been instructive in the study of plant cell differentiation and morphogenesis. The role of trichomes is thought to be in reducing heat and water loss from the plant by increasing the boundary layer between the epidermal tissue and the environment. Trichomes are also hypothesised to prevent insect attack in some species, either acting as a physical barrier, or by secreting chemicals that deter insects (Oppenheimer *et al.*, 1992; Walker and Marks, 2000).

Like cotton fibres, the trichomes of *Arabidopsis* are single cells, whilst those of other plant species, such as tobacco and snapdragon, are multicellular structures, typically consisting of four or more cells. *Arabidopsis* trichomes originate from specialised epidermal cells, which can be first identified by their enlarged nuclei and increased cell size (Larkin *et al.*, 1997). *Arabidopsis* leaf epidermal cells can differentiate into at least three different cell types, including trichomes, guard cells and trichome accessory cells (Oppenheimer *et al.*, 1992). Trichomes begin to form on the surface of leaves very early in leaf development, when the epidermis is still rapidly dividing. Once the trichome cell and its nucleus have begun to enlarge, elongation takes place in a direction perpendicular to the plane of the leaf. When the developing trichome cell has reached about ten times the size of a regular epidermal cell, branches begin to form and the nucleus moves up to the branching point during maturation when the trichome reaches 300- 500 μm in length (Koornneef, 1981). The first trichomes form near the distal end of the leaf and along the margins, with subsequent trichomes developing in the proximal region (Hülskamp *et al.*, 1994). In a mature leaf, trichomes are distributed evenly over the leaf surface, rarely initiating from adjacent epidermal cells (Figure 1.2B). In *Arabidopsis*, trichomes are found on various parts of the plant, including the rosette, the stem, and the sepals of the flower. The stem trichomes are generally unbranched (Larkin *et al.*, 1997).

Cell differentiation of *Arabidopsis* trichomes has been studied extensively for several reasons. Firstly, trichomes develop on the leaf surface and are easily observed and analysed. Secondly, trichome formation is a relatively simple process compared with other plant cell differentiation events and thirdly, trichomes are not required for experimental plant growth and development, so it is possible to view the effects of mutations that alter trichome development without deleterious effects on the plant (reviewed by Schellmann and Hülskamp, 2005).

1.8 Genetic regulation of Arabidopsis trichome initiation

The identification of mutants defective in trichome development has been made easier by the fact that trichomes are not essential for *Arabidopsis* survival under experimental conditions. Most of these mutant loci can be divided into two classes, based upon plant phenotype (Hülskamp *et al.*, 1994). The first class of genes includes those which are

involved in the initiation of trichome development. Mutants in this class alter the initiation frequency and spacing of trichomes and these are of most interest in this study. The second class contains those genes involved in trichome cell expansion and morphogenesis. Defects at these loci do not affect trichome initiation but result in trichomes with abnormal morphology, such as uncharacteristic cell shape and branching.

1.8.1 TRANSPARENT TESTA GLABRA1 (TTG1)

One gene that is known to regulate trichome initiation in *Arabidopsis* is *TRANSPARENT TESTA GLABRA1* (*TTG1*). Loss-of-function *ttg1* mutations result in an almost complete absence of leaf trichomes (Figure 1.2C; Koornneef, 1981). The *TTG1* gene is pleiotropic and *ttg1* mutants also lack anthocyanins in the leaf and hypocotyl. *TTG1* has been shown to regulate numerous steps of the flavonoid biosynthesis pathway in *Arabidopsis* (Figure 1.3). The mucilage normally exuded by the seed coat is absent in *ttg1* mutants (Koornneef, 1981) which also display altered root hair patterning (Galway *et al.*, 1994). A number of *ttg* alleles have been characterised, with each affecting the full range of mutant phenotypes to varying degrees (Koornneef, 1981; Larkin *et al.*, 1994; Larkin *et al.*, 1997). The observation that each allele displayed the complete set of phenotypes provided the first indication that *TTG1* encodes a protein with a single function required in several pathways, as opposed to containing multiple functional domains affecting the different *TTG1* functions. A small number of trichomes develop in weak *ttg* alleles (such as *ttg1-9* and *ttg1-10*). However, these alleles also demonstrate a high incidence of clustered trichomes in adjacent cells, in contrast to the even spacing observed in wild-type, suggesting a role for *TTG1* in regulating the spatial distribution of trichomes.

The *TTG1* locus has been mapped to chromosome 5 of *Arabidopsis*, and it encodes a polypeptide of 341 amino acid residues containing four WD-40 repeats (Walker *et al.*, 1999). The WD-40 motif contains a 44-60 residue sequence, usually with a GH (Gly-His) dipeptide 11-24 residues from its N-terminus and a WD (Trp-Asp) at the C-terminus and a conserved core between these dipeptides (Smith *et al.*, 1999). WD-40 repeats have been identified in a number of proteins from a wide variety of eukaryotic organisms and they occur in large gene families. There are 136 identified family members in humans, 175 in *Drosophila melanogaster*, 106 in *C. elegans*, 60 in *S. cerevisiae* and 59 in *Arabidopsis* (Yu *et al.*, 2000a; Li and Roberts, 2001, http://bmerc-www.bu.edu/cgi-bin/wd_db). WD-40 proteins have been implicated in processes such as signal transduction, RNA processing,

gene regulation, vesicular traffic, protein interaction and regulation of cell cycle (Neer *et al.*, 1994), although none are thought to be an enzyme. Over 30 subfamilies among the WD-repeat-containing proteins have been identified, the majority of which correlate with distinct functions or cellular roles. Their common function appears to be co-ordination of the formation of multiprotein complexes, mediated by the repeating units, which serve as a scaffold for protein-protein interactions. The identification of TTG1 as a WD-40 repeat protein indicates that, rather than functioning directly as a transcription factor, TTG1 interacts with other proteins in playing its various roles.

TTG1 is a member of the *transparent testa* (*tt*) gene family, a group of genes involved in the flavonoid biosynthesis pathway, which provides plants with a variety of secondary metabolites such as anthocyanins (red to purple pigments), flavonols (colourless to pale yellow pigments) and condensed tannins (colourless pigments that brown with oxidation - also known as proanthocyanins) (Figure 1.3). Anthocyanins accumulate in vegetative tissue, condensed tannins are deposited in seeds, whilst flavonols are observed in both vegetative tissue and seeds (reviewed in Winkel-Shirley, 2001). Flavonoids play many roles in plant development, including protection against microbial attack and the effects of UV radiation, signalling agents and regulators of auxin transport. Mutations at the *transparent testa* loci disrupt the synthesis of brown pigments in the seed coat (testa), rendering the seed coat transparent. *tt* mutants therefore produce seeds of a yellow or pale brown colour, as the transparent seed coat reveals the yellow colour of the underlying cotyledons. To date, approximately 20 *transparent testa* loci have been identified in *Arabidopsis*, most of which have been fully characterised. These genes encode either regulatory proteins or proteins involved in the structural properties which are summarised in Table 1.1

TTG1 is expressed in all tissues of *Arabidopsis* (Walker *et al.*, 1999), an expression pattern in common with other WD-repeat genes in *Arabidopsis* and other plant species. The TTG1 protein sequence demonstrates the highest similarity to WD-repeat proteins with presumed roles in anthocyanin biosynthesis in a range of plant species (deVetten *et al.*, 1997; Sompornpailin *et al.*, 2002; Morita *et al.*, 2006). Due to its critical role in trichome initiation, the *TTG1* gene is the major focus of this thesis.

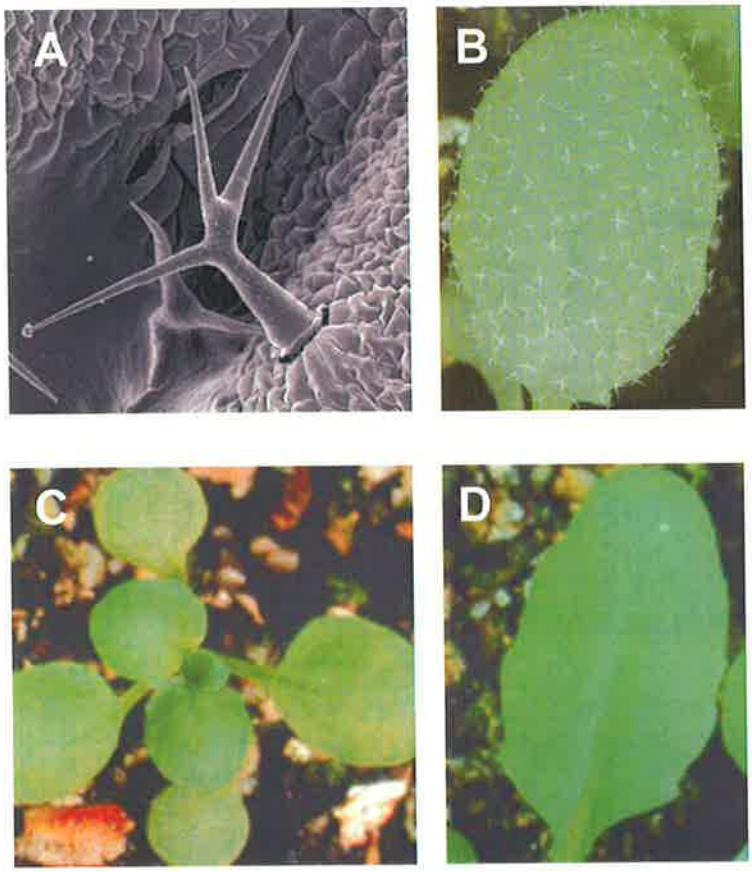


Figure 1.2: *Arabidopsis* trichomes and mutant phenotypes

(A) SEM image of a three-branched *Arabidopsis* trichome. **(B)** Trichomes developing on a wild-type *Arabidopsis* leaf. **(C, D)** Glabrous leaves of *ttg1* (C) and *gl1* (D) mutant *Arabidopsis*. Images B-D from Larkin *et al.* (1994).

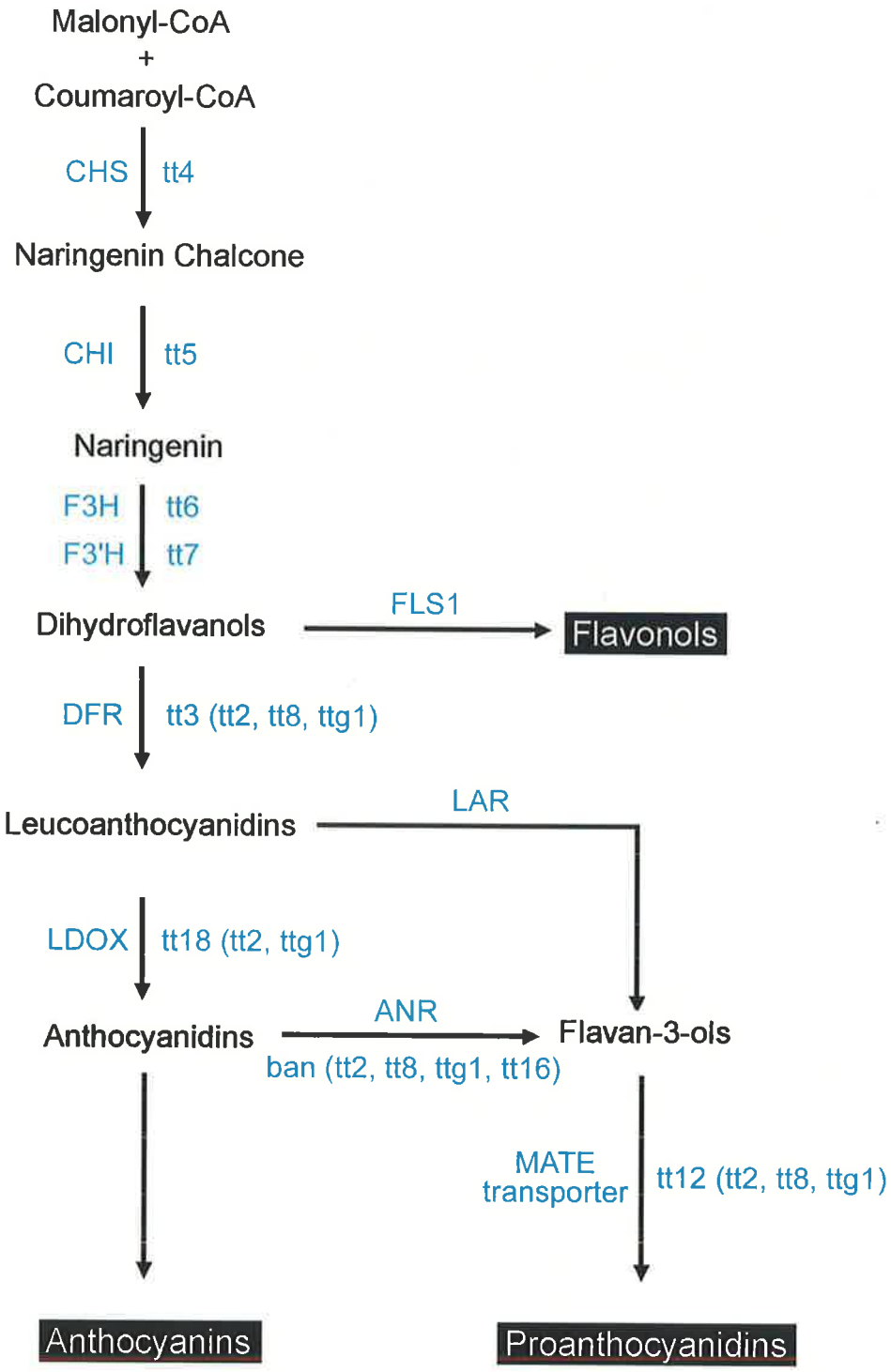


Figure 1.3: Flavonoid biosynthesis pathway in *Arabidopsis*

Enzymatic steps affected in specific *transparent testa* mutants are indicated, with regulatory genes indicated in brackets. CHS = chalcone synthase; CHI = chalcone flavanone isomerase; F3H = flavanone 3-hydroxylase; F3'H = flavonoid 3'-hydroxylase; DFR = dihydroflavonol 4-reductase; FLS1 = flavonol synthase; LDOX = leucocyanidin dioxygenase; LAR = leucoanthocyanidin reductase; ANR = anthocyanidin reductase; MATE = multi drug and toxin extrusion.

Locus	Corresponding protein product	Reference
Structural genes		
tt3	Dihydroflavonol 4-reductase	(Shirley *et al.*, 1992)
tt4	Chalcone synthase	(Feinbaum and Ausubel, 1988)
tt5	Chalcone isomerase	(Shirley *et al.*, 1992)
tt6	Flavanone 3-hydroxylase	(Pelletier and Shirley, 1996)
tt7	Flavonoid 3'-hydroxylase	(Koornneef *et al.*, 1982)
tt10	Polyphenol oxydase	(Pourcel *et al.*, 2005)
tt12	Multi-drug secondary transporter	(Debeaujon *et al.*, 2001)
tt15	Glycosyltransferase	(Focks *et al.*, 1999)
tt18	Leucocyanidin dioxygenase	(Shikazono *et al.*, 2003)
tt19	Glutathione S-transferase	(Kitamura *et al.*, 2004)
Regulatory Genes		
tt1	Zinc-finger containing protein	(Sagasser *et al.*, 2002)
tt2	R2R3 MYB domain protein	(Nesi *et al.*, 2001)
tt8	bHLH protein	(Nesi *et al.*, 2000)
tt16	MADS domain protein	(Nesi *et al.*, 2002)
ttg1	WD40 protein	(Walker *et al.*, 1999)
ttg2	WRKY protein	(Eulgem *et al.*, 2000)
Other		
tt9	Unknown	(Shirley *et al.*, 1995)
tt13	Unknown	(Debeaujon *et al.*, 2000)
tt17	Unknown	(Bharti and Khurana, 2003)

Table 1.1: Members in the transparent testa gene family, with their corresponding protein product (where known) and appropriate reference.

1.8.2 GLABROUS1 (GL1)

Like *TTG1*, *GLABROUS1* (*GL1*) is a required gene in the initiation of *Arabidopsis* trichomes, with *gl1* mutants being essentially hairless (Figure 1.2D, Marks and Feldmann, 1989). MYB transcription factors make up a very large family of proteins in *Arabidopsis*, with the majority (including GL1) containing two MYB repeats (R2 and R3). The family of R2R3-type *MYB* genes are known to control many aspects of plant secondary metabolism, as well as plant cell identity and fate (Stracke *et al.*, 2001). In general, the R2R3 MYB proteins contain a transcriptional activation domain at the C-terminus, in addition to MYB domains that function in DNA binding and protein interactions. *GL1* has been shown to be expressed predominantly in the early stages of trichome development (Larkin *et al.*, 1993) and, unlike TTG1, its role appears to be restricted to trichome

formation. A similar MYB gene (*MYB23*) exists in *Arabidopsis* that is functionally equivalent to *GL1* in trichome initiation on leaf edges, which plays an additional role in trichome branching (Kirik *et al.*, 2005).

1.8.3 GLABRA3 (GL3)

It has been shown that the maize *R* gene, which encodes a basic helix-loop-helix (bHLH) transcription factor, is able to rescue the full range of mutant phenotypes of *Arabidopsis ttg1* (Lloyd *et al.*, 1992). This observation led to the initial suggestion that *TTG1* might encode an *Arabidopsis* R homologue. However, the identification of TTG1 as a WD-repeat protein led to the hypothesis that TTG1 may instead regulate or interact with a bHLH protein in its role in trichome initiation. Payne *et al.* (2000) isolated the *GLABRA3* (*GL3*) gene, which encodes a bHLH protein associated with trichome formation. The effects of *gl3* mutations are less severe than ~~that~~ of *ttg1* and *gl1*; *gl3* plants are not devoid of trichomes, but produce fewer trichomes than wild-type, and those trichomes that do form are not fully developed. It was demonstrated via a yeast two-hybrid assay that GL3 physically interacts with both GL1 and TTG1, although GL1 and TTG1 do not interact directly with each other (Payne *et al.*, 2000). It is thought that the MYB GL1 protein binds to the N-terminal end of the GL3 protein, whilst TTG1 appears to interact at the C-terminal end, which contains the bHLH motif (Grotewold *et al.*, 2000; Payne *et al.*, 2000). It was therefore hypothesised that an activator complex consisting of TTG1-GL3-GL1 was responsible for promoting trichome initiation in *Arabidopsis*.

1.8.4 TRANSPARENT TESTA8 (TT8) and ENHANCER OF GL3 (EGL3)

Over-expression of *GL3* suppresses the glabrous phenotype in *ttg1* plants, but does not influence other *TTG1*-related traits. The fact that *gl3* mutations result in only a moderate reduction in trichome number, and that *gl3* mutants do not have an obvious effect in any other TTG1-dependent pathways suggested that other partially redundant bHLH proteins may be responsible for the remainder of these pathways, given that the *R* gene was able to complement the full range of *ttg1* mutant phenotypes. The *TRANSPARENT TESTA8* (*TT8*) gene, which is involved in flavonoid biosynthesis in *Arabidopsis*, codes for a bHLH protein (Nesi *et al.*, 2000) but, whilst mutations in this gene confer the transparent testa phenotype due to a decrease in condensed tannin production, *tt8* plants still produce significant amounts of anthocyanin. This indicates that at least one additional bHLH

protein is likely to act in anthocyanin production, trichome initiation and other TTG1-controlled pathways.

A genetic enhancer screen in the *gl3-1* background identified a locus that resulted in glabrous plants, when combined with *gl3-1* (Zhang *et al.*, 2003). This locus contained a close relative of *GL3*, and was termed *ENHANCER OF GL3* (*EGL3*). *EGL3* also encodes a bHLH protein, which shares 74% amino acid identity with GL3. Like GL3, the EGL3 and TT8 bHLH proteins have been demonstrated to interact with the TTG1 protein in yeast two-hybrid assays (Zhang *et al.*, 2003; Baudry *et al.*, 2004). The *egl3/gl3* double mutant appears similar to the *ttg1* mutant, displaying reduced seed coat mucilage, altered root hair positioning and disrupted anthocyanin production. A triple bHLH mutant (*egl3/gl3/tt8*) essentially replicates the *ttg1-1* mutation, displaying the complete set of mutant phenotypes. This demonstrates that the three bHLH proteins have distinct but overlapping functions in TTG1-dependent pathways, and explains how the maize R gene was able to suppress the entire range of *ttg1* mutant phenotypes (Figure 1.4).

1.8.5 TRIPTYCHON (TRY) and CAPRICE (CPC)

The *TRIPTYCHON* (*TRY*) and *CAPRICE* (*CPC*) genes have been identified as playing a role in trichome initiation in *Arabidopsis*, most likely involved in cell-cell signalling (Hülskamp *et al.*, 1994; Wada *et al.*, 1997). The *TRY* and *CPC* genes are partially redundant, and each has been shown to encode related single repeat (R3) MYB-type transcription factors that function in lateral inhibition (Wada *et al.*, 1997; Schellmann *et al.*, 2002). Developing trichomes are distributed in an even spacing pattern; trichome cells are on average separated by about four epidermal cells at the point of initiation in wild-type *Arabidopsis* leaves, indicating that lateral inhibition of differentiation of neighbouring cells is occurring (Hülskamp *et al.*, 1994). Mutations in the *TRY* or *CPC* genes increase the number of trichome clusters observed, demonstrating a decrease in nearest-neighbour inhibition (Hülskamp *et al.*, 1994; Schnittger *et al.*, 1999; Schellmann *et al.*, 2002). *TRY* and *CPC* also have been shown to be partially redundant with two genes which enhance their mutant phenotype; *ENHANCER OF TRY AND CPC 1* and *2* (*ETC1* and *ETC2*) (Kirik *et al.*, 2004a; Kirik *et al.*, 2004b). Each of the encoded single-repeat MYB proteins are thought to act via cell-to-cell movement to provide competition with R2R3 MYBs (such as GL1) for GL3/EGL3 binding site interactions in neighbouring cells. The single-MYB

proteins lack a transactivation domain, and thus form an inactive complex when bound to GL3/EGL3 (Figure 1.5).

1.8.6 GLABRA2 (GL2)

The *GLABRA2* (*GL2*) gene influences trichome morphogenesis in *Arabidopsis*. *GL2* function is not required for initiation; *gl2* leaves exhibit a normal number of trichomes but is thought to function early in the trichome development pathway (Szymanski *et al.*, 1998). *GL2* encodes a homeodomain protein and mutations result in aborted trichomes with aberrant cell expansion (Rerie *et al.*, 1994). Trichomes in *gl2* mutants expand laterally along the surface of the leaf, as opposed to normal cell outgrowth perpendicular to the leaf surface. Szymanski *et al.*, (1998) demonstrated that *GL2* is expressed in trichome cells throughout development, as well as in surrounding cells. *GL2* positively regulates the trichome cell fate in the shoot, and non-hair cell fate in the root. Transcription of *GL2* is regulated by the TTG1-GL3-GL1 activation complex in trichome initiation, and by a related complex in root hairs (see Section 1.8.7, and Figure 1.5).

1.8.7 Trichome and root hair initiation

The protein complexes involved in trichome initiation are similar to those involved in the patterning of root hairs. However, in contrast to the processes controlling leaf trichome patterning, *Arabidopsis* root hair cells are specified in a position-dependent manner, as only those cell files that are positioned above the junction of two underlying cortical cells produce root hairs (Dolan *et al.*, 1994). The cells located between two underlying cortex cells (Figure 1.5) adopt the root-hair fate, whilst those positioned above only one cortical cell develop as non-hair cells. The TTG1 protein in *Arabidopsis* is one of several which regulate root hair patterning (Galway *et al.*, 1994). In common with trichome patterning, a complex is formed involving the TTG1 protein, an R2R3 MYB-type protein, and EGL3/GL3. In the case of root hairs, the R2R3 MYB-type protein involved in the activator complex is WEREWOLF (WER), the functional homologue of GL1. The expression of the TTG1, GL3/EGL3 and GL1 or WER activator complex in the shoot or root respectively promotes the expression of GL2, which allows for specification of the trichome cell fate in leaves, and the non-hair cell fate in the root. The activator complex also up-regulates the expression of the single repeat MYBs TRY/CPC/ETC1/ETC2, which move through the plasmodesmata to mediate lateral repression in neighbouring cells. It is hypothesised that the R3 MYB proteins interfere with the formation of the bHLH-TTG1-R2R3MYB

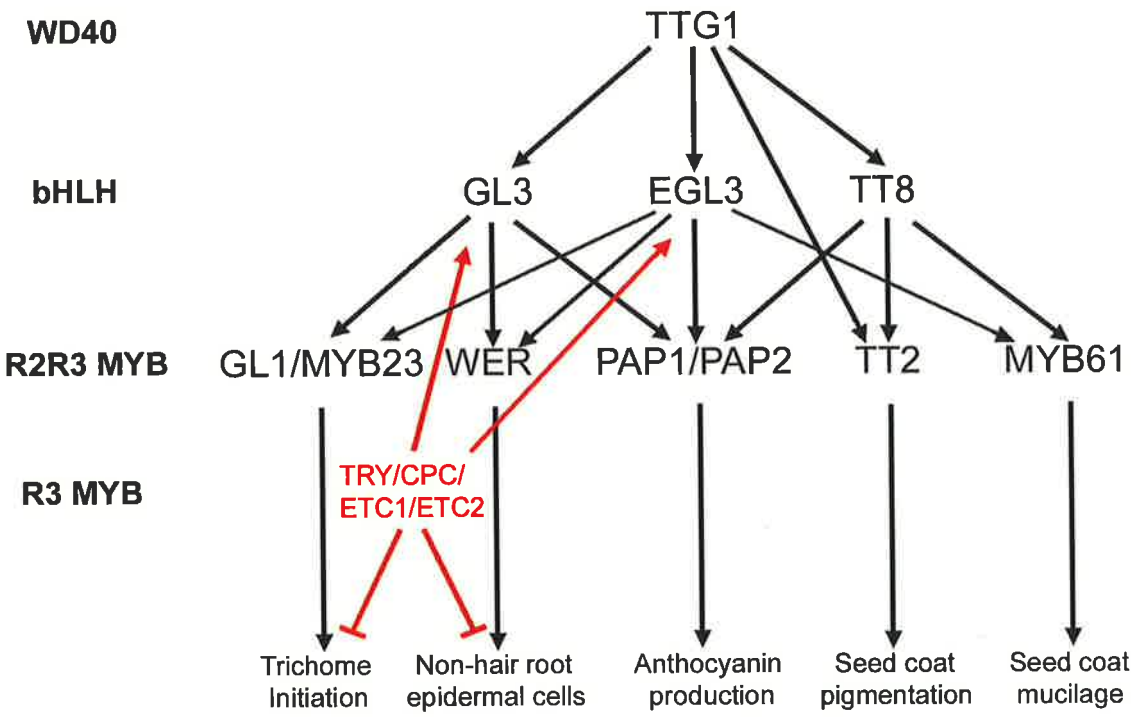


Figure 1.4: Model of interactions involved in TTG1-related processes

This model shows all of the known bHLH and MYB transcriptional regulators that function in the TTG1-dependent pathways. The model illustrates the interactions among the proteins, the narrowing of the specificity of the regulatory function from WD40 to bHLH to MYBs, and the redundancies demonstrated, particularly at the bHLH level. Arrows indicate known interactions, whilst the red blocked lines denote inhibition of the particular cell fate. Adapted from Zhang *et al.* (2003).

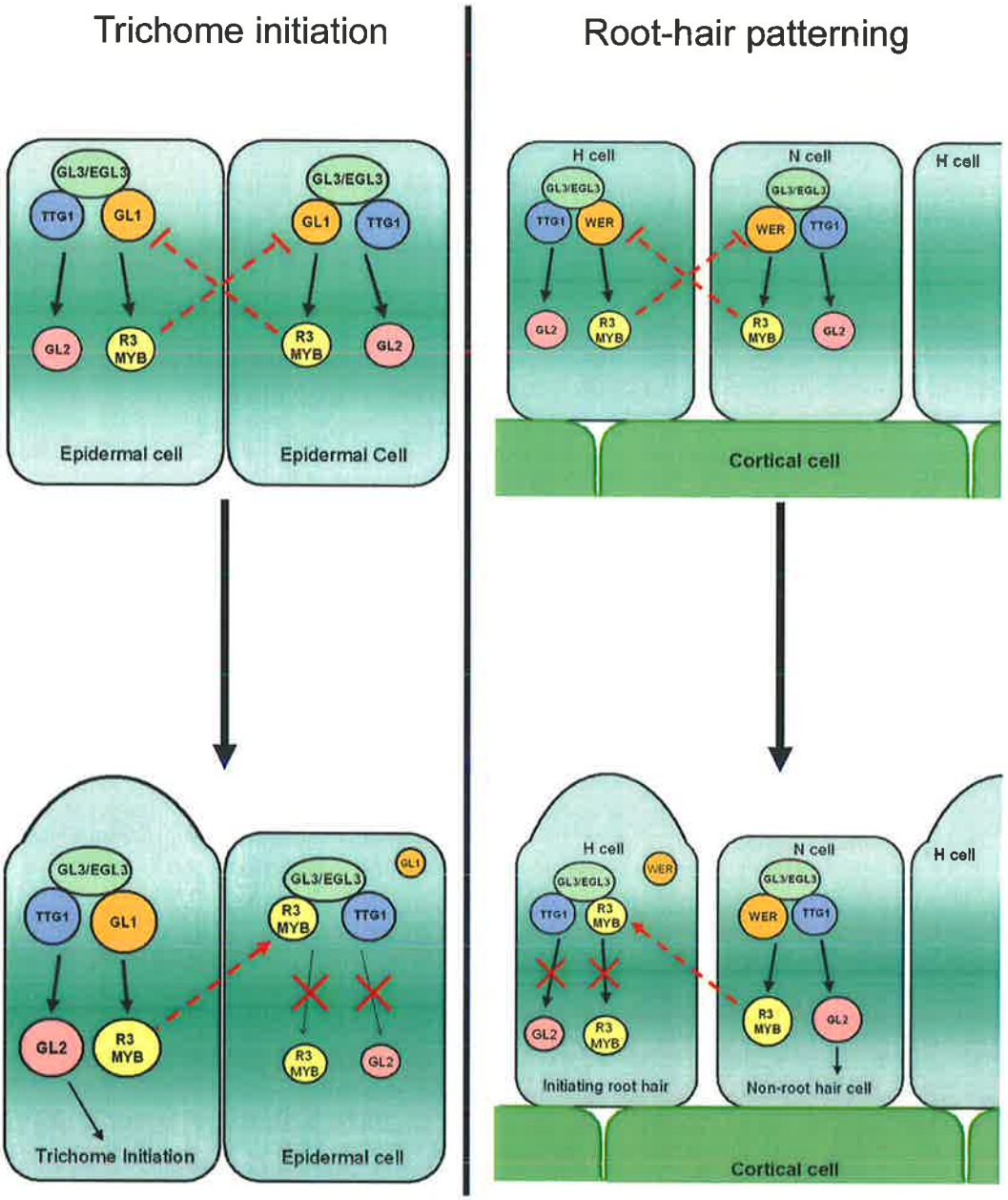


Figure 1.5: Model for initiation of trichome (left) and root hair patterning (right) in *Arabidopsis*

The upper tier shows cells with equivalent developmental potentials, whilst the lower tier shows cells that have been specified to adopt a particular fate. In the case of trichome initiation, a cell receives an unknown signal, resulting in up-regulation of the TTG1-GL3-GL1 complex, leading to activation of GL2 transcription and adoption of the trichome cell fate. The single MYB protein (TRY/CPC/ETC1/ETC2; represented by R3 MYB) moves to neighbouring cells to form the inactive complex, resulting in the non-trichome fate in these cells. It is a similar case in the root, except it is the position of the underlying cortical cells that determines epidermal cell fate. A root epidermal cell in the H position (between two underlying cortical cells) adopts the hair cell fate, whereas a cell in the N position (above a single cortical cell) adopts the non-hair cell fate.

complex, promoting the adoption of the hair cell fate in roots, or the non-trichome fate in the leaf (reviewed in Schellmann and Hülskamp, 2005). Recent evidence suggests GL3/EGL3 may also move between cells as part of the regulatory network in root hairs, although further verification is required (Schiefelbein and Lee, 2006).

1.8.8 The WD40/bHLH/MYB regulatory complex

As discussed, the function of TTG1 in a range of processes involves interplay with various MYB and bHLH transcription factors. The WD40-bHLH-MYB complex is conserved throughout several plant species, and is involved in multiple cellular differentiation pathways in the plant epidermis (reviewed in Ramsay and Glover, 2005). The complex controls epidermal cell identity and regulates distribution of specialised epidermal cells. Although its individual components are more ancient, this complex appears to have arisen in the land plant lineage, and is estimated to have evolved at least 120 million years ago (Ramsay and Glover, 2005). One of the reasons for the extensive use of this complex in plants is its flexibility and plasticity; it is an example of an established network producing new phenotypic traits by variation in the MYB and, to a lesser extent, the bHLH components. Whilst significant variation is observed within these components, the multiple protein-protein interactions occurring do place some constraints on the evolution of the system. For example, the maintenance of the seven-bladed β-propeller structure of the WD40 repeat protein is critical for the stability of the complex, and the multiple interactions in which the bHLH is involved constrains its variability.

As detailed in Figure 1.4, TTG1 interacts with numerous bHLH and MYB proteins to influence a wide range of functions. In its role in flavonoid production in *Arabidopsis*, TTG1 has been shown to interact with TT8 (a bHLH factor) and TT2 (an R2R3 MYB domain protein). In the latter case, a direct interaction has been shown between TT2 and TTG1, and this complex is able to activate expression of the *BANYULS* gene, which encodes a core enzyme required for seed coat pigmentation (Baudry *et al.*, 2004). In addition, several R2R3 MYB proteins have been shown to be involved in TTG1-related functions in *Arabidopsis* by interacting with bHLH proteins, such as PRODUCTION OF ANTHOCYANIN PIGMENT1 (PAP1) and PRODUCTION OF ANTHOCYANIN PIGMENT2 (PAP2) in anthocyanin biosynthesis (Zhang *et al.*, 2003; Teng *et al.*, 2005) and MYB61 in seed coat mucilage production (Penfield *et al.*, 2001).

With the exception of some of these TTG1-related processes in *Arabidopsis*, it is largely in anthocyanin production that a WD40/bHLH/MYB complex has been shown to be active. In addition, there are several cases where MYB and bHLH transcription factors have been shown to act cooperatively in anthocyanin production, without a WD40 protein being identified (reviewed in Grotewold *et al.*, 2000; Elomaa *et al.*, 2003; Sweeney *et al.*, 2006). It could be that a WD40 protein is not playing this role in certain species, or that one simply has not yet been isolated. Examples of WD40/bHLH/MYB complexes functioning in the anthocyanin biosynthesis pathway of various species are detailed below.

The bHLH R protein (which complements the *ttg1* mutant in *Arabidopsis*) has been to shown to interact with the MYB-type protein C1 to activate flavonoid biosynthesis in maize (Goff *et al.*, 1992). Furthermore, the *pale aleurone color1* (*pac1*) gene in maize encodes a WD-repeat protein that interacts with R in its role in anthocyanin pigmentation in the aleurone (Carey *et al.*, 2004). A WD40/bHLH/MYB complex is also thought to be involved in floral pigmentation in *Petunia hybrida*, where ANTHOCYANIN11 (AN11, a WD-repeat protein), ANTHOCYANIN1 (AN1, a bHLH protein) and ANTHOCYANIN2 (AN2, a MYB protein) function in controlling the transcription of dihydroflavanol-4-reductase (DFR) in the anthocyanin biosynthesis pathway (deVetten *et al.*, 1997; Spelt *et al.*, 2002). In *Perilla frutenscens*, Sompornpailin *et al.*, (2002) have shown via a yeast two-hybrid assay, that a WD-repeat protein (PFWD) involved in anthocyanin production, interacts with a member of the *Perilla* bHLH family, which in turn interacts with a MYB-type transcription factor. PFWD does not interact directly with the MYB protein. A study of genes associated with anthocyanin biosynthesis in Japanese morning glory identified a MYB (InMYB1) and WD-repeat protein (InWDR1), which are hypothesised to act coordinately with a bHLH transcription factor (InbHLH2) to activate genes for anthocyanin biosynthesis (Morita *et al.*, 2006).

1.9 Arabidopsis trichomes as a model for cotton fibre development

Arabidopsis leaf trichomes are a potential model for uncovering the genetic mechanism controlling cotton fibre differentiation, as the developmental processes in both appear to be very similar. Both structures arise from the developing epidermis, giving rise to single-celled hairs. In each process, development proceeds across the epidermal surface in a wave

and the initiation stages are very similar, in that the cell expands outwards, and the nucleus enlarges.

It is currently unknown how closely trichomes from different plant species are related, and some data suggest that the pathways of trichome differentiation differ between *Arabidopsis* and other dicotyledons, such as tobacco, snapdragon and petunia (Walker and Marks, 2000). Payne *et al.*, (1999) gave an indication that trichomes of different species may be controlled by distinct pathways. In this study over-expression of GL1 in tobacco had no effect on trichome development, in contrast to the increase in trichome number observed in *Arabidopsis*. Similarly, expression of the *MIXTA* MYB gene from snapdragon and *CotMYBA* from cotton had no effect in *Arabidopsis*, but increased trichome numbers in tobacco. However, phylogenetic analyses in angiosperms have demonstrated that cotton is more closely related to *Arabidopsis* (both being members of the Rostids) than these other dicot species of the Asterid sub-family (Soltis *et al.*, 1999), and thus the mechanism of trichome development in cotton may be more closely related to that of *Arabidopsis*.

One difference that has been described between the two systems is that *Arabidopsis* trichomes undergo several rounds of endoreduplication, whilst cotton fibres are not believed to undergo such a process (Taliercio *et al.*, 2005). However, recent investigations suggest there is some level of endoreduplication occurring in the fibre cell, although not to the level observed in *Arabidopsis* trichomes (Wu *et al.*, 2006).

Several previous experiments have been performed on the basis of analogy between cotton fibres and *Arabidopsis* leaf trichomes. In an attempt to isolate *GL1* homologues in cotton, Loguerico *et al.*, (1999) identified six MYB genes, although none were found to have a role in fibre development. Suo *et al.* (2003) used a degenerate PCR approach to identify a further 55 cotton *MYB* genes, including one (termed *GhMYB109*) which displayed 51% identity to GL1 (including 82% identity within the R2R3 MYB domain) with expression patterns suggestive of a role in fibre development. Another cotton *MYB* gene, *GaMYB2*, was identified and found to be expressed specifically in fibre cells, and to complement the phenotype of *gl1 Arabidopsis* (Wang *et al.*, 2004). Additionally, in the cDNA microarray analysis by Wu *et al.* (2006) to identify genes with expression profiles consistent with a role in fibre initiation, cDNAs with similarity to genes known to play a role in *Arabidopsis*

trichome development (*MYB* genes and a homeodomain gene) were among those identified.

Hsu *et al.*, (1999) demonstrated a connection between the development of cotton fibres and tobacco trichomes. In this case, the promoter activity of a fibre-specific cotton lipid transfer protein gene, *Ltp6*, was analysed. Fused with a GUS reporter, the *Ltp6* promoter was introduced into tobacco leaf discs and shown to direct GUS expression specifically in tobacco trichomes. Similarly, Liu *et al.* (2000) and Delaney (2005) have shown that cotton promoters can direct trichome-specific GUS expression in transgenic tobacco, suggesting cotton fibres and tobacco trichomes share some similarities.

Taken together, these findings indicate that similar regulatory mechanisms may be operating between the trichomes of some species, and that *Arabidopsis* leaf trichomes represent an appropriate model for the study of cotton fibre differentiation. Given these similarities, it is possible that the genes regulating *Arabidopsis* trichome initiation, such as *TTG1*, may have homologues in cotton that function in fibre initiation. The isolation of genes regulating fibre initiation, and hence fibre yield and quality, is essential for improved genetic manipulation of cotton fibre properties.

1.10 Experimental Approach

The primary aim of this project is to identify and characterise genes involved in the initiation of cotton fibres. This will provide information that may be useful in manipulating crop quality and fibre yield. For this purpose, homologues of the *TTG1* gene have been sought in cotton, owing to the role of this gene in *Arabidopsis* trichome initiation. Four putative homologues of *Arabidopsis TTG1* were isolated previously in this laboratory in cotton, termed *GhTTG1-4* (Humphries *et al.*, 2005), identified due to their sequence similarity to *TTG1*.

In the investigation described in this thesis, the genomic organisation of the *GhTTG* genes was analysed via Southern blot and PCR analysis. This was performed to establish whether any of the closely related *GhTTG* genes represent alloalleles, and to gain further insight into the organisation of the WD-repeat gene family in cotton.

It was previously demonstrated that two of the four *GhTTG* genes are able to replace the function of *TTG1* by complementing the anthocyanin-deficient phenotype of a *ttg1* mutant of *Matthiola incana* (Humphries, 2001). To further investigate the similarity of the *GhTTG* genes to *TTG1*, a functional analysis was carried out in *Arabidopsis*. *Agrobacterium*-mediated transformation of an *Arabidopsis ttg1* mutant by the cotton *GhTTG* genes was performed to determine if they are able to complement the glabrous and the other pleiotropic phenotypes conferred by the *ttg1* mutation.

To investigate the expression of the putative *TTG1* homologues in cotton, RT-PCR was performed on fibres and other cotton tissues. *In situ* hybridisation was performed on sectioned cotton ovules at multiple time points to analyse the spatial distribution of *GhTTG* mRNA accumulation, with particular attention being paid to the ovule epidermal cells destined to become fibre cells.

As discussed, TTG1 in *Arabidopsis* controls trichome initiation through protein-protein interactions. Potential interactors of the GhTTG3 protein in cotton were isolated via a yeast two-hybrid screen of a fibre cDNA library. This investigation provides information on which to build a model of proteins that potentially may be required for initiation of cotton fibre cells, based on those involved in initiation of the model *Arabidopsis* trichome system.

Chapter 2: Materials and Methods

2.1 Plant material

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

Gossypium hirsutum cultivar Siokra 1-4

G. raimondii

G. herbaceum

G. hirsutum cultivar Xu-142

G. hirsutum lintless lines: (provided by Yingru Wu, CSIRO).

Xu-142*fl*

Lintless 5B

Lintless 53

Lintless 1A

SL1-7-1

Lintless 4A

Arabidopsis thaliana ttg1-1 seeds were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC), NASC ID# N89

2.2 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
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

2.3 Radioactive isotopes

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

2.4 Molecular weight markers

λ DNA restricted with *Hin*dIII	MBI Fermentas
2-Log DNA ladder	NEB

2.5 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-α-gal (5-Bromo-4-chloro-3-indolyl-a-D-galactopyranoside)	BD Biosciences

2.6 Kits

GenElute Plasmid Miniprep Kit	Sigma-Aldrich
pGEM-T Easy	Promega
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
QIAquick Plasmid Midiprep Kit	Qiagen
Matchmaker Library construction & Screening Kit	BD Biosciences
YEASTMAKER Yeast Transformation System 2	BD Biosciences
DNeasy Plant Mini Kit	Qiagen
DIG RNA Labelling Kit (SP6/T7)	Roche

2.7 Stains and dyes

Bromophenol blue	Sigma-Aldrich
Ethidium bromide	Sigma-Aldrich

2.8 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:

3-Amino-1,2,4-triazole (3-AT)	Sigma-Aldrich
Antifoam A emulsion	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
Histoclear II	National Diagnostics
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
Triton X-100	Sigma-Aldrich

2.9 Vectors

pGEM-T Easy	Promega
pBluescript SK(-)	Stratagene
pGBKT7	BD Biosciences
pGADT7-Rec	BD Biosciences
pART7	Gleave, 1992
pART27	Gleave, 1992

2.10 Yeast and bacterial strains

Species and strain	Source	Use	Genotype and Reference
E. coli DH5αF'	BD Clontech (Palo Alto, CA)	Host for general cloning	*recA1, endA1, gyrA96, thi-1, hsdR17,* (r_K-, m_K+), *supE44, relA1, deoR, Δ (lac ZYA- argF) U169* [F', *ϕ80d lacZ ΔM15*]. (Hanahan, 1983)
A. tumefaciens AGL-1	Dr. John Harvey (CSIRO Division of Plant Industry, Adelaide, Australia)	Plant transformation	AGL0 (C58 pTiBo542) *recA::bla*, T-region deleted Mop^+ Cb^r (Lazo *et al.*, 1991).
S. cerevisiae AH109	BD Clontech (Palo Alto, CA)	Yeast two-hybrid screen	*MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4D, gal80D,* (James *et al.*, 1996)

2.11 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
SP6	GATTTAGGTGACACTTAG
T7	AATACGACTCACTATAG
M13F	GTAAAACGACGGCCAGT
M13R	GTTTTCCGAGTCACGAC
RSP	AACAGCTATGACCATG
Two-hybrid primers	
5AtTTGNde2	GTGATACATATGCTCATTTCCATG
5AtTTGNde4	GTGATTCATATGTCAGCTCCAGATTCG

3AtTTGPST	AACTGCAGACCTCAAACTCTAAGGAGC
5TTG1Nde	GTCTTGCATATGTCAACTCAGGAATCCC
3TTG1Bam	CTGGATCCGAATACATTTCAAGCAG
5TTG2Nde2	GTGATACATATGAGCAGCGATCCTAACCC
3TTG2BAM	CCGGATCCTAGCATTGAATACGACCC
5TTG3Nde3	GTACTGCATATGTCAACTCAAGAATCCC
3TTG3PST	AACTGCAGGACCAGACTTCAAGCAGG
5TTG4Eco	AGCGAATTCACCAGCGATCCGAACCC
3TTG4Pst	TCCTGCAGAACCTAACTTCACACGG
5GL3Eco	GTGAATTCATGCAACTCAGAATCGCC
5GL3Eco2	AGCGAATTCGGACAAAACAGAACAACTGTCC
3GL3Bam	GCGGATCCCCAGCGATCGGCG
5EGL3Eco	AGTGAATTCGGAGAAAACAGAACGGTGCC
3EGL3Bam	GCGGATCCTTAACATATCCATGCAACCC
pGADT7A	CTCCATAATCTTGAAGAGGC
pGADT7B	GGGTACAGGTAGTGTTAGACC
Clone1.5a	GAGATTCTATTCCTGC
Clone1.3a	TGCATGCTGAGATATGTATGG
Clone2.5a	TTCATGTGTGTGGATGTGG
Clone2.3a	GTTCAGCATCCAGTCACC
Clone3.5a	GAAGAATGGTGATGCTGG
Southern Blot primers	
TTG1PRO5	GCGGATCCCCTGCTTGAAATGTATTCAACTCTC
TTG1PRO3	GCGGATCCCTGTCCATATTCCTTGTCTGCTACG
TTG3PRO	GCGGATCCTTGAGGTTTTCACACTGTATTTTCCTGC
TTG4PRO5	AGGTTTGCCAGGCTTCAAGG
TTG4PRO3	ATATACTCCGCAGTCCGCAC
GhTTG gene-specific primers	
PTTG1	GACGGTTGGCAACTCCCAAATAAG
PTTG2	CGCTTCATCCTCTTCTTCTTGTTC
PTTG3	TCCCCTTTCTCTTCCTCCTCCGAC
PTTG5	AGTGTCCGCCGCGACAAGAAATACCGTC
PTTG9	GAGGAGTGCTCCTTGTCGCGCAGG

PTTG10	CTCCGCAACTCAACCACCGGCAAC
PTTG11	TACCGACTCGCCATCGCCAGCCTC
Real-time PCR primers	
TTG1F	CTCTTCCTCCTCCGACCTCCT
TTG1R	CAGAGACGGAGGTAGTCACCG
TTG3F	AGACCCACCCAAAGCTTGC
TTG3R	TGAGCTTGGTGGGTGGATAAG
ActinF	GAGTATCGTTGGTACCCTACC
ActinR	GAGGAGTGCCTTGTCGCGCAGG

2.12 Solutions, buffers and media

All solutions and buffers were prepared using Millipore-filtered water and (when appropriate) were sterilized by autoclaving. Non-autoclavable solutions were sterilized 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:

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

General

Denaturing solution	0.4 M NaOH
Denhardt's solution (1x)	0.02% (w/v) Ficoll, 0.02% (w/v) PVP, 0.02% (w/v) gelatin
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°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)
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)

2.13 Media

All media were prepared with Millipore water and were autoclaved for 15 minutes at 121°C. Non-autoclavable medium components were filter-sterilized 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.

SD agar (-Trp, -Leu ,-Leu/-Trp or QDO)	SD medium with appropriate Dropout supplement and 2% (w/v) agar
SD medium (-Trp, -Leu, -Leu/-Trp or QDO)	26.7g/l Minimal SD Base (BD Clontech, Palo Alto, CA; Cat. #630411)
YEB	1.33% (w/v) nutrient broth, 1% (w/v) yeast extract, 5% (w/v) sucrose, 1 mM $MgSO_4.7H_2O$

YEB agar	YEB medium with 0.9% (w/v) agar
YPD	50g/l YPD Medium (BD Clontech, Palo Alto, CA; Cat. #630409)
YPD agar	YPD medium with 2% (w/v) 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

2.14 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 2-4 per 8-litre pot. Plants were fertilised periodically with commercial fertilisers.

Arabidopsis thaliana ttg1-1 (L*er*) *mutant* seeds were mixed with sand and sprinkled onto "*Arabidopsis* soil mix" (SARDI, Glen Osmond, South Australia) in 3-inch pots. Plantlets were grown at 23°C (light/dark cycle of 8/16 hr) and thinned to 5-10 plants per pot at 2-3 weeks. Plants were allowed to grow for several weeks before transferal to flower-inducing conditions (23°C, light/dark cycle of 18/6 hr). The first flowering bolts (10 cm tall) were cut back once, and new bolts allowed to emerge for several days before infiltration.

Agrobacterium-mediated transformation of *Arabidopsis*

The expression cartridges of the *TTG1*, *GhTTG1*, *GhTTG2*, *GhTTG3* and *GhTTG4* pART7 constructs were removed by *Not*I restriction and inserted into similarly-cut binary vector, pART27 (Gleave, 1992). Plasmid DNA was used to transform *Agrobacterium tumefaciens* AGL-1 (Lazo *et al.*, 1991) by electroporation (Walkerpeach and Velten, 1994) and

colonies selected on plates containing kanamycin. Mutant *ttg1-1* plants were transformed by floral dip infiltration (Clough and Bent, 1998) using a suspension of *Agrobacterium* at OD_{600} = 0.8 in 1.5 % (w/v) sucrose and 0.05 %(w/v) Silwet L-77. Plants were covered overnight to maintain humidity and grown for four weeks to produce seed. Selection of transformants was conducted on 0.8% agar containing Murashige and Skoog (MS) salts (2.2 g/L) and kanamycin (35 μg/ml) for seven days. Between 8,000 and 16,000 seeds were screened for each experiment, with transformation rates varying between 0% and 0.5%. Kanamycin-resistant seedlings were moved onto fresh kanamycin plates for ten days before transfer to soil to set seed, and were inspected microscopically for restoration of trichome formation. Progeny from self-fertilised primary transformants were examined for complementation of other *ttg1* mutant phenotypes. Three-day old seedlings grown on 0.8% agar containing MS salts and 5% sucrose were inspected for anthocyanin production in cotyledons and hypocotyls. Root hair development was studied in seedlings grown vertically on plates containing MS salts and 1.2% phytagel. Seeds for scanning electron microscope (SEM) analysis of seed coat structure were mounted, sputter-coated with carbon and gold, and examined using a Philips XL20 SEM at 10kV. Ruthenium red staining of seed coat mucilage was performed as described by Burn *et al.* (2002).

2.15 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).

Arabidopsis genomic DNA was isolated from the leaves of putative transformants using the DNeasy Plant Mini Kit (Qiagen).

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 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 a polytron was used in place of a glass tissue grinder. RNA was precipitated and stored in 0.1 volumes of 3M sodium acetate and 2 volumes of 100% ethanol at -70°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 nm, or by gel electrophoresis of a sample and comparison with *Hin*dIII-digested λ DNA fragments.

Restriction endonuclease digests

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.

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 by 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 redissolved in 1x TE.

DNA precipitation by 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 redissolved in 1x TE.

Gel electrophoresis

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

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 manufacturer'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), essentially as described in the manufacturer's protocol with the modification of using half the described amount of reaction mix. 400 – 800 ng of double-stranded DNA was used as a 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 8 and later (Accelrys, San Diego, CA; Genetics Computer Corporation, WI, 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 Center for Biotechnology Information http://www.ncbi.nlm.nih.gov).

2.16 Southern blot analysis

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.

Southern transfer to membranes

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.

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 adding the labelled samples in stop buffer in a total volume of 100 µl, to the top of the column and re-centrifugation at 170 *g* for two minutes.

Prehybridisation conditions

All filters were prehybridised in a freshly made solution of hybridisation buffer at 65°C for a minimum of two 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

Radiolabeled 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 ml of hybridisation solution (to a maximum volume of 20 ml) for a minimum of 16 hours. All hybridisations were performed at 65°C overnight.

Washing of filters

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 20 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

Filters were covered in plastic wrap to prevent them from drying out and exposed to X-ray film at varying temperatures (-70°C, -20°C, 4°C or room temperature) for an appropriate period of time, backed by an intensifying screen (DuPont Hi-Plus). Typically, genomic Southerns were exposed for at least four days at -70°C.

2.17 In situ hybridisation

Paraffin embedded sections

Paraffin-embedded sections of cotton ovules were prepared as described by Houben *et al.* (2005). Briefly, cotton ovules were collected at 0 DPA and 2 DPA and fixed in formalin-acetic acid. The fixed ovules were then subjected to a series of tertiary butyl alcohol washes, and embedded in paraffin. 8 μm sections of ovules were cut using a microtome, and placed onto slides in preparation for addition of the probe.

In vitro transcription reactions

GhTTG2 and *GhTTG3* sequences were cloned into the pGEM-T Easy vector, and each clone was digested downstream of the insert to linearise the plasmid, and creating a 5' overhang. *In vitro* transcription reactions were performed as described in the DIG RNA Labelling kit protocol (Roche), with the T7 and SP6 RNA polymerase being used to prepare sense and anti-sense RNA (summarised in Table 5.1). A 1 kb segment of the cotton ribosomal RNA gene was used as a positive control cloned into pGEM-T Easy for generation of a positive control probe.

Hybridisation and detection of probe

In situ hybridisation of the paraffin-embedded sections and subsequent immunological detection was carried out according to the procedure described by Marrison and Leech (1994).

2.18 Yeast two-hybrid assay and yeast methods

The yeast two-hybrid assay was performed using the MatchMaker cDNA Library Construction and Screening Kit supplied by BD Biosciences. Yeast culture techniques, transformation and x-α-galactosidase assays were performed as described in the Two-Hybrid system manual (Clontech Part No. PT1031-1) and the Yeast Protocols Handbook

(Clontech Part No. PT3024-1). Library screening was performed on QDO media, supplemented with 5mM 3-AT.

Plasmid DNA was isolated from yeast cultures using the method of Hoffman and Winston (1987), with the exception that yeast cells were ground in a MM300 mixer mill (Qiagen) for two and a half minutes at 30Hz instead of vortexing. Also, the crude plasmid DNA extract was concentrated by adding 2.5 volumes of ethanol and 1.3% (v/v) 7.5M ammonium acetate; placing at -80°C for one hour; centrifuging for ten minutes at 13,000 *g*; drying the pellet in a vacuum centrifuge and resuspending in 20 µl of water. Transformation of *E. coli* was then performed with 10 µl of the concentrated solution. Plasmid DNA in *E. coli* was examined by restriction digests and sequence analysis.

Chapter 3: WD-repeat genes in cotton

3.1 Previous work: Isolation of four WD-repeat genes in cotton with homology to TTG1

The *TTG1* gene plays a critical role in trichome initiation in *Arabidopsis*. Given the potential for trichomes of *Arabidopsis* to be used as a model system for cotton fibre development, genes in the cotton genome with homology to *TTG1* have been sought, with the view they may play a role in the initiation of the cotton fibre. In work performed previously in this laboratory (see Humphries *et al.*, 2005), a combination of library screening and PCR-based techniques was used to isolate putative homologues of *Arabidopsis TTG1* from cotton. Degenerate *TTG1* PCR primers were used to amplify a 700 bp fragment from cotton genomic DNA. Nucleotide sequencing of four cloned amplicons revealed two distinct, but similar, sequences. Probing of a 12 DPA cotton fibre cDNA library with one of the inserts detected two cDNA clones, one of which, GhTTG1, corresponded exactly in sequence to the genomic PCR clone used as probe (also called GhTTG1). The other cDNA, GhTTG2, was a new *TTG1*-like isolate. The second genomic PCR sequence was designated GhTTG3. Probing a cotton genomic DNA library with genomic *GhTTG1* identified three additional clones, one of which corresponded exactly in sequence to GhTTG1, whilst the remaining two clones, which overlapped with each other, contained a fourth *TTG1*-like gene, *GhTTG4*. The *GhTTG2* and *GhTTG3* genes were obtained using PCR-based genome walking, providing genomic clones for all four genes.

In the cases of *GhTTG1* and *GhTTG2*, only minor sequence differences were observed between the gene and a corresponding cDNA, and it was concluded that both genes and their corresponding transcripts had been cloned. The sequence of the *GhTTG1* gene was partial, with the subcloning site located 2 bp downstream from the presumed initiation codon. cDNA sequences were therefore used in all sequence comparisons which involved GhTTG1 (1253 bp) and GhTTG2 (1694 bp), and gene sequences were used in the cases of *GhTTG3* (2008 bp) and *GhTTG4* (1837 bp). Comparison between the gene and cDNA sequences showed that all four genes lack introns in their protein coding regions, as is the case for *Arabidopsis TTG1* (Walker *et al.*, 1999) and other plant WD-repeat genes (deVetten *et al.*, 1997; Sompornpailin *et al.*, 2002; Carey *et al.*, 2004). The *GhTTG3* sequence overlapped completely with a cDNA termed ghttg1 (AF336281), also isolated

from ovules of *G. hirsutum* cv. Acala Maxxa on the day of flowering (Matz and Burr, unpublished). Since only three (non-synonymous) single nucleotide substitutions were observed in the 1341 bp of overlap, *GhTTG3* is almost certainly the gene which is transcribed to give the ghttg1 mRNA.

Sequence comparisons between the four deduced proteins and *Arabidopsis* TTG1 showed that they form two groups, with GhTTG1 (343 amino acids) and GhTTG3 (345 amino acids) being closely related to each other (87% identical and 93% similar) and to TTG1 (79% and 80% amino acid identity respectively). GhTTG2 and GhTTG4 (both 346 amino acids) formed the second group, with 95% amino acid identity to each other and only 62-63% identity to TTG1 (Table 3.1).

	AtTTG1	**GhTTG1**	**GhTTG2**	**GhTTG3**
GhTTG1	79% (89%)	-		
GhTTG2	63% (79%)	62% (79%)	-	
GhTTG3	80% (89%)	87% (93%)	63% (78%)	-
GhTTG4	62% (79%)	60% (78%)	95% (97%)	61% (77%)

Table 3.1: Amino acid identity, with similarity in brackets, between cotton and *Arabidopsis* TTG1 sequences

Analysis of the four amino acid sequences showed that GhTTG1-4 are basic proteins with molecular weights of 38-39 kDa, similar to TTG1. Also, like TTG1, they appear to have a seven-bladed repeat structure with four clearly defined WD repeats (Yu *et al.*, 2000a), as indicated in Figure 3.1, and are therefore members of the large and diverse WD-repeat protein family. The sequence similarity between the proteins extends across the WD-repeats and to the carboxyl terminus (Figure 3.1).

Database searches using the four sequences confirmed the GhTTG1/GhTTG3 and GhTTG2/GhTTG4 groupings, and identified similarities to WD-repeat proteins from cotton and other plants. The GhTTG1 and GhTTG3 protein sequences both showed a high degree of similarity to WD-repeat proteins which have known or implied roles in anthocyanin biosynthesis, with 79-80% amino acid identity to *Arabidopsis* TTG1 (Table 1; Walker *et al.*, 1999), 79% amino acid identity to AN11 (de Vetten *et al.*, 1997), 82%

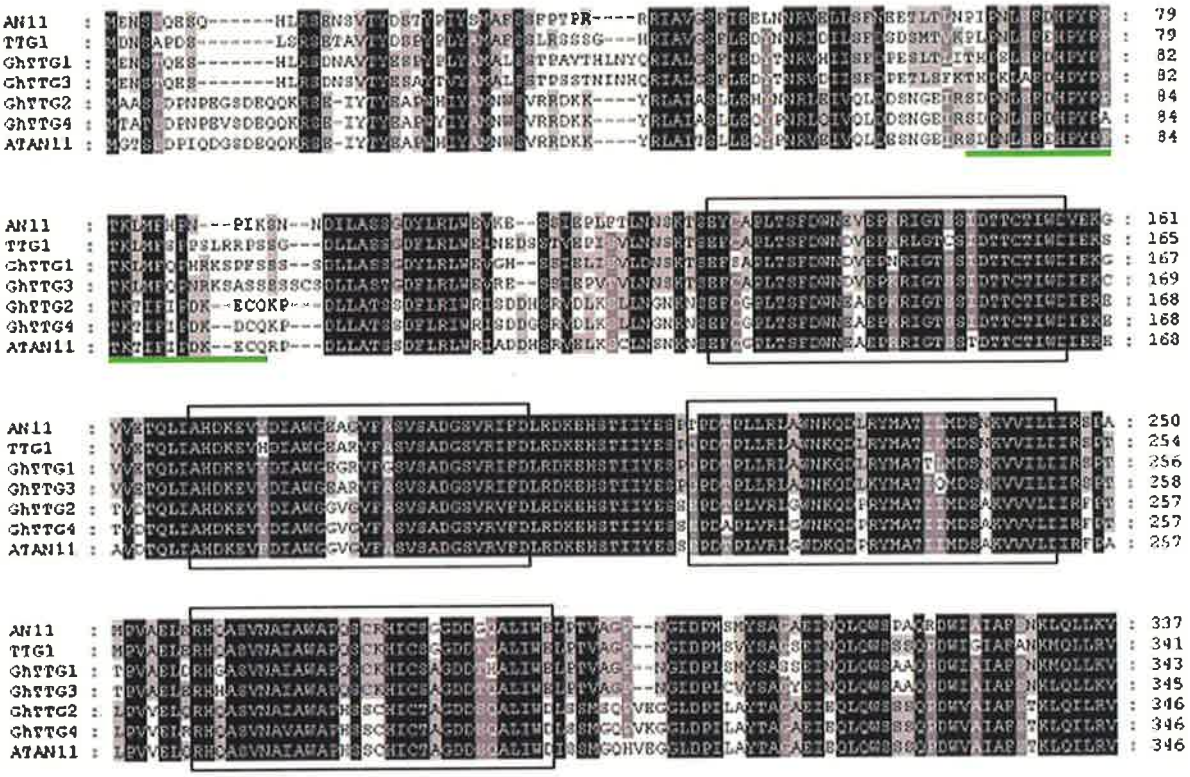


Figure 3.1: Deduced amino acid sequences of four WD40 proteins isolated from cotton, compared with *Arabidopsis* TTG1 and ATAN11-A and petunia AN11

Dark shading indicates positions where all seven sequences are identical or chemically conserved, while lighter shading shows positions where there is a lower level of conservation. The positions of the WD repeats, as determined by the BMERC Protein Structure Prediction Server (http://bmerc-www.bu.edu/) are indicated with boxes. The proline-rich region is underlined in green.

identity to the deduced product of a *TTG1*-like gene from apple (Casas-Mollano and Destefano-Beltrán, 2000) and 77-78% to PFWD from *Perilla frutescens* (Sompornpailin *et al.*, 2002). GhTTG2 was virtually identical to a 1401 bp cDNA, wd1522 (AF336287), isolated from ovules of *G. hirsutum* cv. Acala Maxxa on the day of flowering (Matz and Burr, unpublished). The protein sequences differed by one residue, though the conceptual translation of GhTTG2 was 32 amino acids longer than that of wd1522, due to different choice of start codons. The GhTTG2 and GhTTG4 deduced protein sequences both showed a high degree of similarity to *Arabidopsis* ATAN11-A (de Vetten *et al.*, 1997), with 92% and 90% amino acid identity, respectively. High similarity has also been observed to WD-repeat proteins from two varieties of the flowering plant Morning Glory (*Ipomoea nil* and *Ipomoea purpurea*), termed InWDR2 and IpWDR2 respectively (Morita *et al.*, 2006). Lesser similarity was observed to the maize MP1 protein, and a number of uncharacterised, putative WD-repeat-containing proteins revealed by various plant genome sequencing projects (see Section 3.3- Discussion).

The four nucleotide sequences were also virtually identical to cotton ESTs sequenced as part of the International Cotton Genome Initiative. The GhTTG1 cDNA appears to represent a full-length version of an EST sequence (BF271126) from a *Gossypium arboreum* 7-10 DPA fibre cDNA library (Wing *et al.*, unpublished), whilst both *GhTTG2* and *GhTTG3* correspond to ESTs (AI731000 and AI729712, respectively) from a *G. hirsutum* cv. Acala Maxxa 6 DPA fibre cDNA library (Blewitt *et al.*, unpublished). *GhTTG4* appears to represent a full-length version of an EST (DT462698) from a *G. hirsutum* ovule cDNA library (Udall *et al.*, 2006).

In addition to the WD-repeat regions, the *Arabidopsis* TTG1 protein contains a proline-rich region, in which eight out of 26 amino acid residues are proline. In the corresponding region of the GhTTG proteins, six prolines are present with the exception of GhTTG4 (five prolines) (Figure 3.1). It is unknown whether this proline-rich region plays a specific functional role, but is thought to take a part in the formation of the propeller blade structure of WD-repeat proteins (Walker *et al.*, 1999).

3.2 Genomic organisation of cotton WD-repeat genes

3.2.1 Genomic PCR analysis

G. hirsutum is a tetraploid cotton thought to have arisen 1-2 million years ago by hybridisation between a maternal A genome taxon and a paternal D genome taxon (Wendel, 1989). To determine the sub-genomic origin of each cotton *TTG1*-like gene, a PCR-based approach was employed, using genomic DNA isolated from three AD allotetraploid cotton varieties, namely *G. hirsutum* cultivars Siokra 1-2 and Siokra 1-4 and *G. barbadense* cv Pima S-6, and two extant species (*G. raimondii* Ulbr. (D_5 genome) and *G. herbaceum* L. (A_1 genome)) which best represent the putative diploid ancestors.

In order to create a precise assay for each of the four cotton *TTG1*-like genes, primer pairs were designed to specifically amplify each gene (Materials and Methods). These primers were complementary to the most variable region of each gene, outside of the conserved WD-repeat regions.

After optimising parameters including annealing temperature and number of cycles, PCR was performed on all four *TTG1*-like clones to demonstrate that each primer pair resulted in amplification of a single band of predicted size only from template DNA of the expected clone (Figure 3.2). These PCR reactions therefore allow for a specific assay for each of the four cotton genes.

PCR was performed using each of the gene-specific primer pairs on genomic DNA extracted from the three tetraploid species and the two extant diploid species. As expected, each primer pair generated a product of predicted size from the three tetraploid species. However, none of the primer pairs generated a PCR product with the A-genome DNA as template, and all amplified a product from *G. raimondii* DNA (Figure 3.3). This indicates that each of the four cotton *TTG1*-like genes originates from the ancestral D-genome. This result was somewhat surprising, given the high similarity that exists between GhTTG1 and GhTTG3, and between GhTTG2 and GhTTG4. It was originally hypothesised that these genes may represent allo-alleles, with each gene of the closely related pairs originating from a different ancestral diploid genome. However, it appears all four *TTG1*-like genes isolated in this study are derived from the D genome of *G. raimondii*, and therefore do not form two homoeologous pairs. As no product had been amplified from the A-genome sample with any primer set, PCR was performed on the same genomic DNA samples using

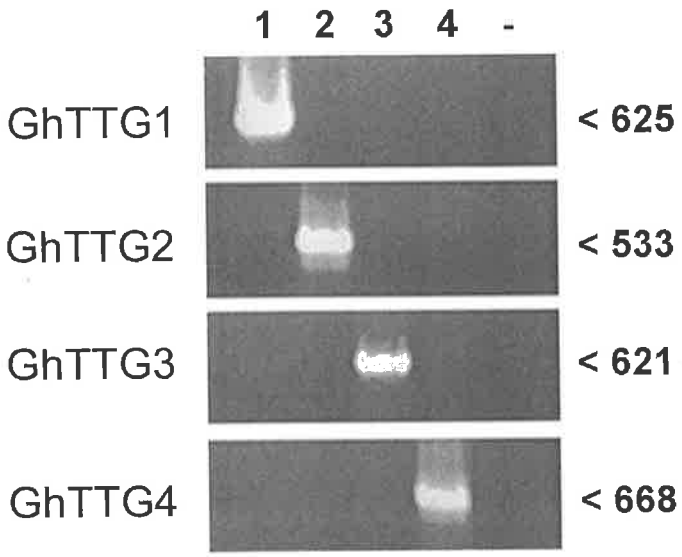


Figure 3.2: Gene-specific PCR assay

PCR reactions of template DNA from each of the clones indicated, with lanes 1-4 representing clones of genes *GhTTG1-GhTTG4*, respectively. Numbers on the right indicate the size of the PCR product in bp. The gene name on the left indicates the target gene of each specific primer pair. For each assay, the specific gene being targeted was the only one to show an amplification product. The 5th lane (-) is a no template control.

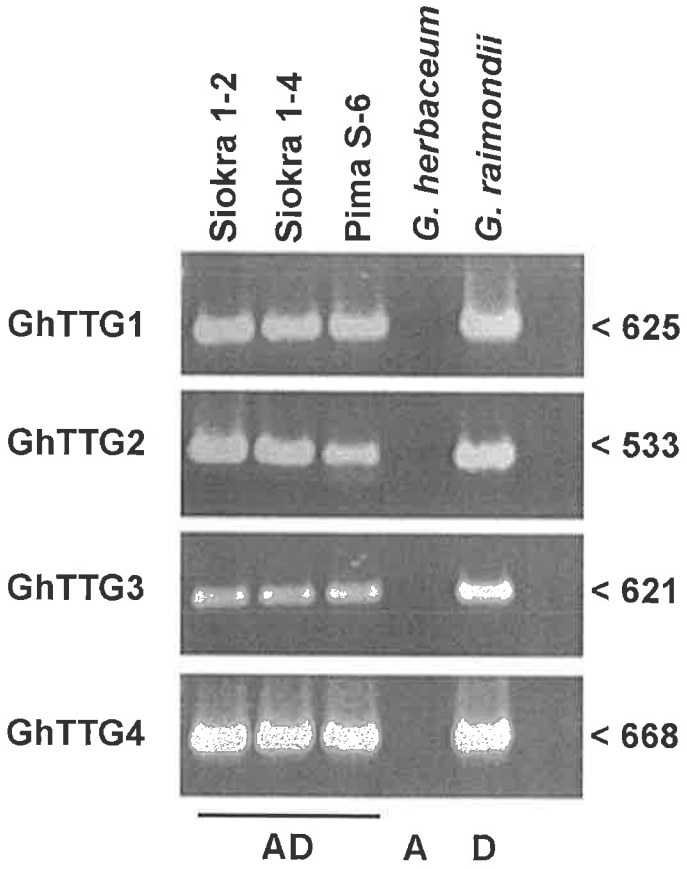


Figure 3.3: Genomic origins of *TTG1*-like genes in cotton

Each lane contains PCR products obtained from genomic DNA templates. Numbers on the right indicate the size of the PCR product in base pairs, and the genomic constitution of each *G. hirsutum* cultivar or diploid species is denoted below the gel.

primers designed to specifically amplify a control gene known to be derived from the A-genome to validate the result. A product was observed in the tetraploid species and A-genome DNA, but not in the D-genome sample, authenticating the results observed with the *GhTTG* primers (result not shown).

3.2.2 Southern blot analysis of WD-repeat genes

To further investigate the genomic organisation of the WD-repeat genes in cotton, the full-length isolates were initially used in turn to probe Southern blots of cotton genomic DNA restricted with various restriction endonucleases. DNA from the diploids *G. herbaceum* and *G. raimondii* were included in the experiment to investigate further the genomic origins of WD-repeat genes in cotton. The genes within each pair (*GhTTG1/GhTTG3* and *GhTTG2/GhTTG4*) cross-hybridised strongly to each other, producing identical Southern hybridisation patterns, and some cross-hybridisation was also observed between the different sets of probes (results not shown). The hybridisation patterns to cotton genomic DNA restricted with a range of restriction enzymes are indicative of only a small number of closely related genes in tetraploid cotton, rather than large gene families. The two sets of probes also hybridised strongly to sequences present in *G. herbaceum* and *G. raimondii* (Figure 3.4A) indicating that, whilst the four particular genes described here are derived from the D diploid genome, the A genome also contains closely related WD-repeat genes.

Additional probes were designed to specifically target each of the cotton WD-repeat genes, in order to obtain hybridisation patterns specific for each gene. This was achieved by utilising the unique 3'UTR of each gene. Primers were designed to amplify by PCR a gene-specific segment of the 3' non-coding region of each gene (see Materials and Methods), which was subsequently labelled and used as a probe in the Southern hybridisation.

For this component of the analysis, each genomic DNA sample (*G. hirsutum, G. raimondii* and *G. herbaceum*) was digested with *Eco*RI, and the *G. hirsutum* DNA was digested with a selection of restriction enzymes. The resulting Southern blots were probed with each of the gene-specific probes in turn, followed by stringent washing.

Each of the Southern blots was characteristic of hybridisation to a specific *TTG1*-like sequence, and showed each of these genes is present as a single or low copy gene in the

cotton genome. Strong hybridisation was observed in the *G. raimondii* (D-genome) lane with each probe (Figure 3.4B). A weaker signal was detected in the A-genome donor, *G. herbaceum*, suggesting the presence of similar but diverged genes. Each probe produced a different hybridisation pattern, indicating the probes were to some extent specific for their intended target. Only a single band was observed in the majority of the lanes and, when two bands were present, a corresponding restriction site generally existed within the gene (or 3'UTR) being probed that was assumed to be maintained in the different taxa. The Southern hybridisation pattern for *GhTTG3* showed two main bands in each lane, perhaps indicating some level of cross-hybridisation with *GhTTG1*, or another related gene. This genomic Southern analysis of the four *GhTTG* genes suggested that there exists a small family of related WD-repeat genes in tetraploid cotton, that each of the GhTTG genes originate from the ancestral D-genome, and is present at a single or low copy number.

3.3 Phylogenetic analysis of cotton WD-repeat sequences

An extensive phylogenetic analysis carried out by Carey *et al.*, (2004) established that two separate groups of related WD-40 proteins exist in plant species. In this study, Genbank, the Plant Genome Database and The Index for Genomic Research (TIGR) were investigated to identify homologues of the *Z. mays* WD-repeat proteins PAC1 and MP1. The multiple alignments of 75 protein sequences suggested there exists at least two classes of the WD40 protein family in plants (Carey *et al.*, 2004). One class has been designated the PAC1 clade which includes PAC1, TTG1 and AN11. This group of proteins includes several with a defined function in anthocyanin biosynthesis. This is in contrast to the second class of WD40 plant proteins, termed the MP1 clade, for which no definite mutant phenotypes have been assigned.

Carey and co-workers included the four cotton protein sequences described in this thesis in their phylogenetic analysis. GhTTG1 (termed Cotton1) and GhTTG3 (Cotton2) grouped with TTG1 in the PAC1 clade, whilst GhTTG2 (Cotton4) and GhTTG4 (Cotton5) grouped with ATAN11A and ATAN11B in the MP1 clade. A fifth cotton WD-repeat sequence (termed Cotton3) appears to fall outside both major clades, but only partial data was provided for this sequence (Carey *et al.*, 2004).

By searching the TIGR and Plant Genome databases, additional sequences were identified in several plant species which show high similarity to the GhTTG1/GhTTG3 group (i.e.

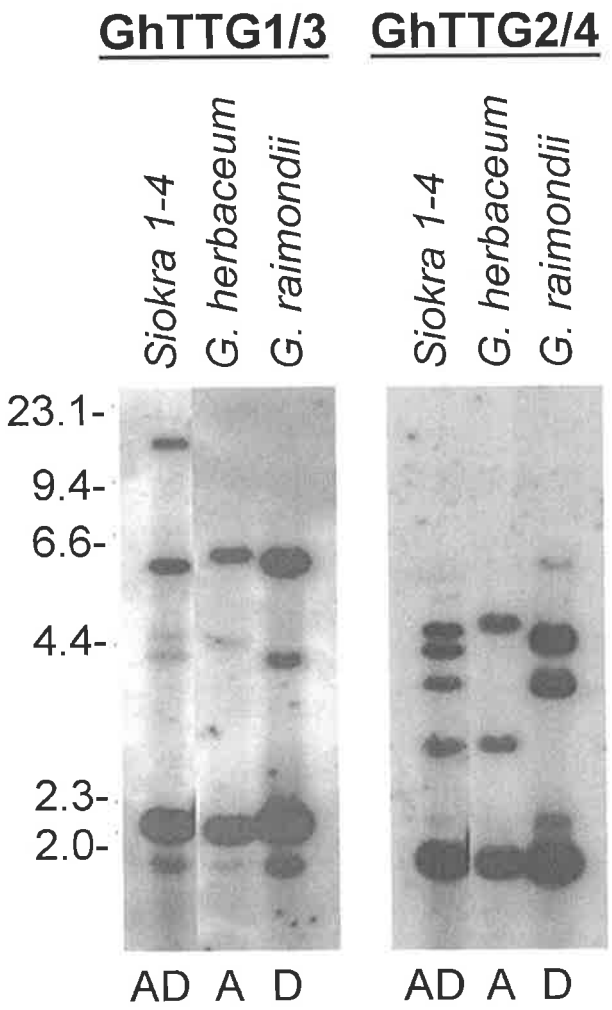


Figure 3.4 (A): Southern analysis of *TTG1*-like genes in tetraploid and diploid cotton

Each lane contains 10 μg of genomic DNA restricted with *Eco*RI and blots are probed with the GhTTG1 cDNA on the left, to represent the results with both *GhTTG1* and *GhTTG3*, and GhTTG2 cDNA on the right, to represent the result with both *GhTTG2* and *GhTTG4*. Numbers on the left indicate molecular weight markers (with sizes indicated in kilobases), and the genomic constitution of each species is denoted beneath each lane.

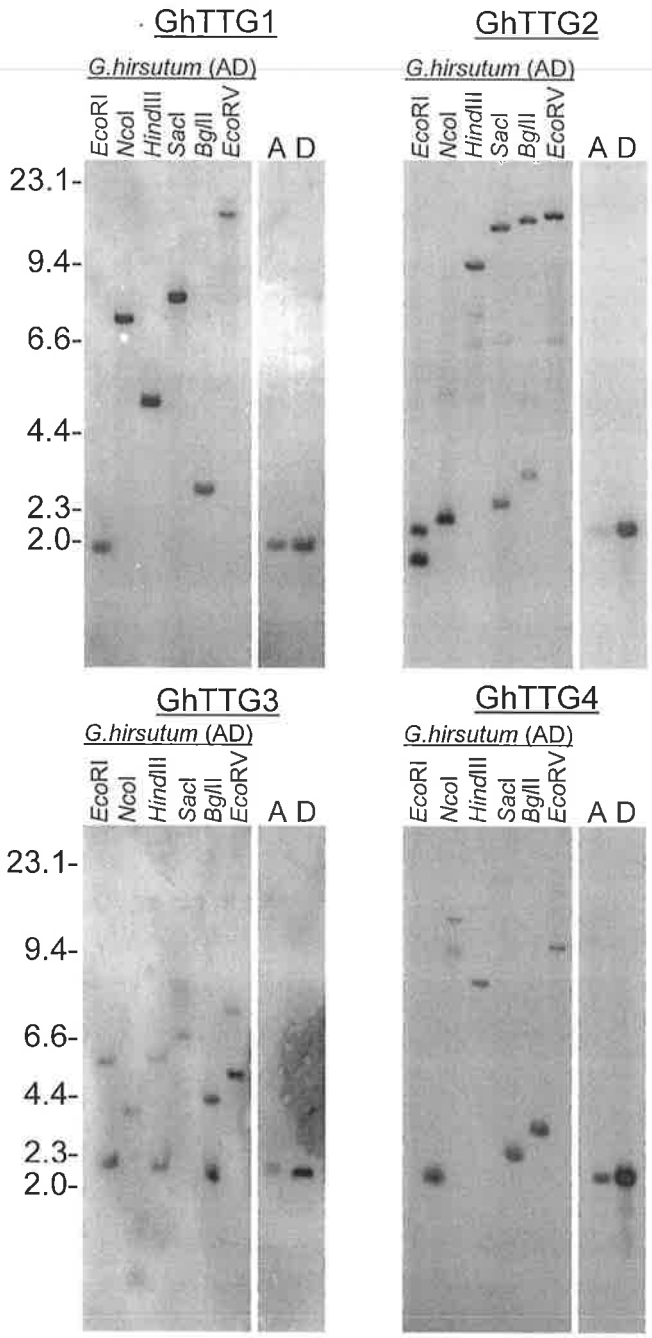


Figure 3.4 (B): Southern analysis of *TTG1*-like genes in tetraploid and diploid cotton using gene-specific probes

Southern blots probed specifically for each of the four cotton *TTG1*-like genes as indicated above each autoradiograph. *G. hirsutum* was restricted with a range of restriction enzymes indicated above each lane, whilst the diploid DNA samples were digested with *Eco*RI. Numbers on the left indicate molecular weight markers with sizes in kilobases, with the genomic constitution of each species denoted above the corresponding lane. The A-genome is represented by *G. herbaceum*, whilst the D-genome sample is *G. raimondii*.

PAC1 clade from Carey *et al.*, 2004) or GhTTG2/GhTTG4 (MP1 clade). Further information can be added to the analysis performed by Carey and co-workers, as more data has become available since this publication. For example, many of the sequences used were partial (around 90 amino acids). For many of these, a sequence that conceptually encodes a full-length WD-repeat protein has now been isolated. After constructing a phylogenetic tree using these full-length sequences, a similar phylogeny to that displayed by Carey *et al.* (2004) is observed (Figure 3.5). The maximum parsimony tree demonstrated that the GhTTG1 and GhTTG3 cotton sequences are closely related to AtTTG1 amongst a range of related WD-repeat sequences.

The phylogeny observed in Figure 3.5 exhibits two major clades (A and B), both of which can be divided into two sub-groups; Clade A contains a distinct PAC1-like group, in addition to a TTG1-like group. Similarly, Clade B contains ATAN11-like and MP1-like sub-groups. The cotton sequences GhTTG1 and GhTTG3 fall into the TTG1-like sub-division, whilst GhTTG2 and GhTTG4 are within the ATAN11-like grouping.

3.4 Discussion

3.4.1 Isolation of four TTG1-like genes in cotton

Four sequences similar to *Arabidopsis TTG1* have been isolated from commercial cotton, based on the notion that they may function in fibre initiation. Analysis of the sequences confirmed that they display high homology to AtTTG1, and each contains four WD-repeat motifs. Smith *et al.* (1999) published a consensus amino acid sequence for the WD-repeat domain, which featured 20 positions (out of total 31-40 residues) that are biased for a particular amino acid (or several amino acids). Of these 20 positions, the WD-repeat regions of the GhTTG proteins are typically identical in 10-11 residues, as is AtTTG1 and several other WD-repeat sequences in the plant database that are involved in anthocyanin production. None of the four cotton TTG1-like proteins contain nuclear localisation signals, suggesting that the proteins reside in the cytosol, or require accessory proteins for import into the nucleus to carry out their presumed role as regulators of transcription.

The high degree of similarity observed between WD-repeat proteins of different species is likely due to high numbers of interacting partners and the pressures of conserving the seven-bladed propeller structure (Neer *et al.*, 1994; Carey *et al.*, 2004). It is commonly observed that WD40 repeat proteins show higher conservation between species than other

proteins in related pathways. For example, when comparing functionally equivalent bHLH and MYB anthocyanin proteins between species, around 40% identity generally exists for bHLH (Payne *et al.*, 2000; Esch *et al.*, 2003) and about 50-60% for MYB sequences (Quattrocchio *et al.*, 1999; Ramsay and Glover, 2005; Quattrocchio *et al.*, 2006). WD-repeat sequences involved in anthocyanin regulation in various species generally exhibit around 80% identity and 90% similarity (de Vetten *et al.*, 1997; Walker *et al.*, 1999; Sompornpailin *et al.*, 2004; Humphries *et al.*, 2005; Ramsay *et al.*, 2005; Morita *et al.*, 2006).

3.4.2 All GhTTG genes originate from the ancestral D genome

An examination of the genomic origins of the *GhTTG* genes suggested that each is derived from the same ancestral diploid genome. The hybridisation patterns observed in the Southern blot analysis appeared to confirm the results from the genomic PCRs, revealing that each of the *GhTTG* genes is derived from the ancestral D genome. This diverged from the hypothesis that *GhTTG* genes may represent allo-alleles, where each gene of the closely related pairs would originate from a different ancestral diploid genome. An example of allo-allelic cotton genes was observed in a recent study of MYB genes in cotton, where two full-length fibre cDNAs were isolated, termed *GhMyb7* and *GhMyb9* (Hsu *et al.*, 2005). The MYB proteins encoded by these genes share approximately 98.5% amino acid identity, and a study of their genomic origins revealed that *GhMyb7* is derived from the A subgenome, whilst *GhMyb9* is from the D subgenome, indicating that the genes are allo-allelic, originating from different ancestral diploid genomes. Similar results have been observed in many other *Gossypium* tetraploid genes studied, including sucrose synthase (Ruan *et al.*, 2003), vacuolar ATPase catalytic subunit (Wilkins and Wan, 1993), acetohydroxyacid synthase (Grula *et al.*, 1995) and tonoplast intrinsic protein (Ferguson *et al.*, 1997). However, such homoeology was not observed for the *GhTTG* genes. Whilst the two pairs of cotton genes are very similar, it has been observed that sequence divergence between the coding regions of homoeologous genes in *Gossypium* is generally in the range of 1%, and that in many cases predicted amino acid sequences are identical (Cronn *et al.*, 1999; Adams *et al.*, 2003; Senchina *et al.*, 2003). In this light, as the amino acid diversion between GhTTG1 and GhTTG3 (13%) and GhTTG2 and GhTTG4 (5%) exceeds this level, it is perhaps not surprising that they are not alloallelic. However, an investigation by Orford and Timmis (2000) found that two fibre-specific LTP genes (*FSltp1* and *FSltp3*) demonstrating 93% amino acid identity are likely to be homoeologous. In this case the

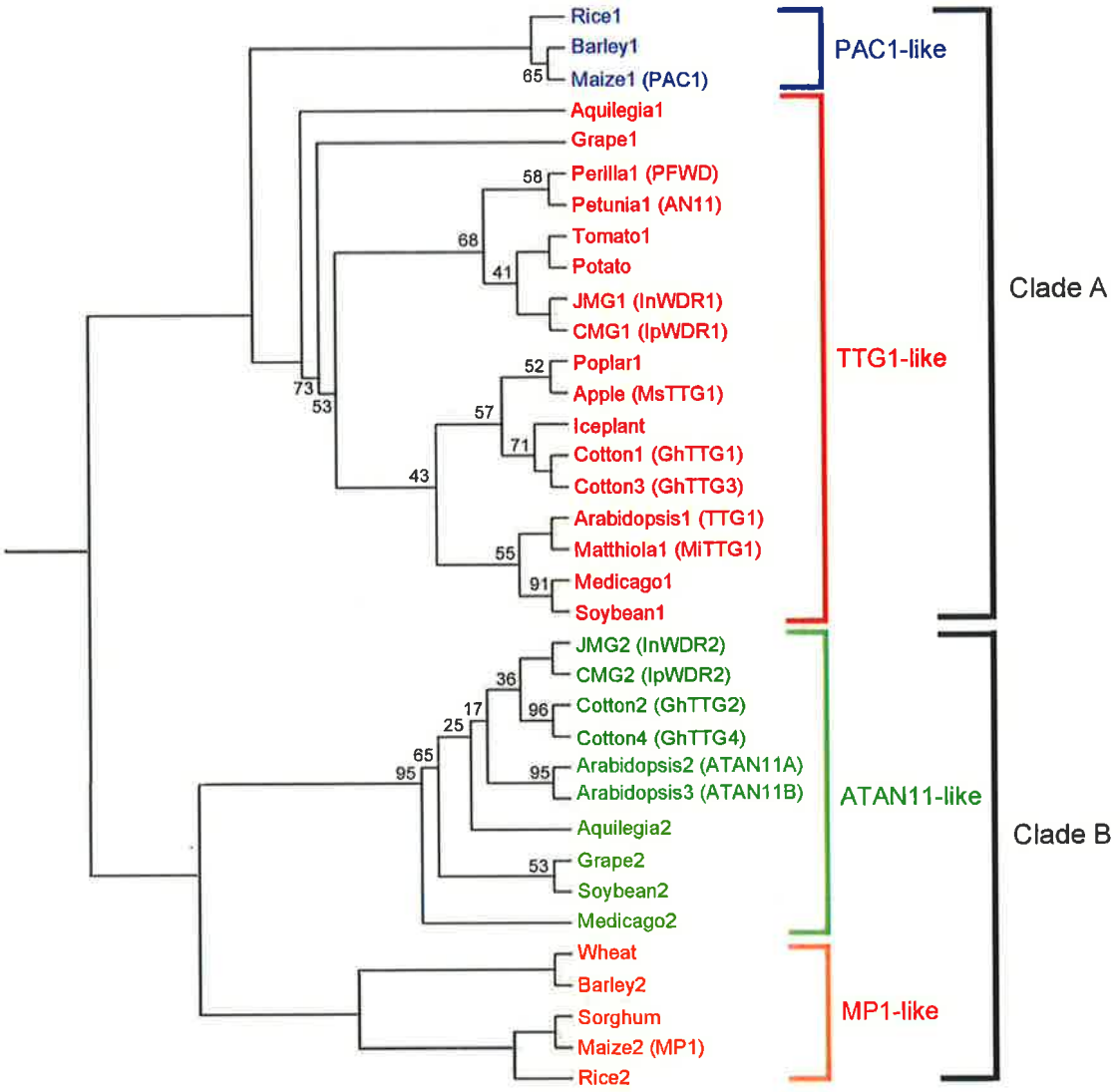


Figure 3.5: Plant WD-repeat genes with similarity to TTG1

Phylogenetic tree of plant WD-repeat sequences, where full-length amino acid sequences were aligned with the CLUSTALW program within the PHYLIP package, and the depicted consensus tree was constructed using the maximum parsimony method on the set of 100 bootstrapped set of protein alignments. Bootstrap values at each branch point are 100 unless indicated. The sub-divisions of the A and B clades are shown in blue (PAC1-like), red (TTG1-like), green (ATAN11-like) and orange (MP1-like). The common names of the plant species from each sequence are shown, with gene names shown in brackets where assigned. JMG= Japanese Morning Glory, CMG= Common Morning Glory. See Appendix 4 for sequence identification numbers.

flanking regions of the genes also demonstrated high homology, which is not the case for the *GhTTG* genes. Given that the pairs of *GhTTG* genes are not alloalleles, the most likely explanation for the high level of nucleotide identity observed is that they are products of a recent gene duplication event, probably occurring after the polyploidisation event hybridising the ancestral A and D genomes about 1 MYA (Wendel, 1989).

It has been shown that the majority of loci which affect fibre quality originate from the D genome (Jiang *et al.*, 1998), with over 70% of the QTLs affecting fibre-related traits falling into D subgenome linkage groups. It would therefore be considered likely that if the *GhTTG* genes were playing a role in fibre development they would derive from the D genome, as this investigation has demonstrated.

3.4.3 Phylogenetic analysis of WD-repeat genes

A previous phylogenetic analysis of WD40 proteins established two separate groups, namely the PAC1 clade and the MP1 clade (Carey *et al.*, 2004). Figure 3.5 represents an updated version of the phylogenetic tree produced by Carey and co-workers (2004), and includes only full-length conceptual sequences. It is evident that the PAC1 clade described by Carey and co-workers (corresponding to Clade A in Figure 3.5) can be divided into two subgroups, namely PAC1-like and TTG1-like. Similarly, Clade B (corresponding to the MP1 clade of Carey *et al.*, 2004) can be divided into an ATAN11-like subgroup and an MP1-like subgroup.

Clade A contains WD-repeat sequences that have been shown to functionally complement the *ttg1* mutant phenotype in *Arabidopsis*, including AN11 and PAC1, as well as PFWD which has demonstrated an effect on *TTG1*-related pathways when over-expressed in *Arabidopsis*. None of the genes of Clade B have been shown to functionally complement *TTG1*, with a negative result being observed for the *mp1* gene (Hernandez, unpublished). Therefore it appears that *GhTTG1* and *GhTTG3* are the most likely of the four cotton genes to display conservation of *TTG1* function, as these genes fall within Clade A.

Genes of Clade A affect several different pathways, with anthocyanin production being in common with all genes for which a phenotype has been described. It thus appears likely that the cotton *GhTTG* genes, in particular *GhTTG1* and *GhTTG3*, will play some role in anthocyanin biosynthesis in cotton, in addition to any possible roles in trichome or fibre

initiation. A question arising from the variety of pathways affected by this clade of proteins is whether the diversification of other (non-anthocyanin) pathways is due to evolution of protein sequences within this group, or the different regulatory networks present within each species (Carey *et al.*, 2004). The fact that genes such as PAC1 and AN11 are able to complement *ttg1* mutant phenotypes in *Arabidopsis* suggests that the difference in pathways affected by these proteins in different species is the result of variations in the mechanism of regulation of these traits, rather than differences at the protein sequence level. The phenotypes that each WD-repeat gene affects could be related to gene duplication events in a given species, resulting in functionally redundant WD40 proteins within the same organism.

Proteins from each of the two major plant WD-repeat clades (Clade A and B) arising from an ancient duplication may still share common functions. It is thought that ATAN11A and ATAN11B from *Arabidopsis* and MP1 from maize may play some role in anthocyanin biosynthesis. However, the fact that several pathways are affected in *ttg1* mutants despite the presence of ATAN11A and ATAN11B indicates that common functionality is not retained, at least in some cases. Thus, it is possible, but not certain that the *GhTTG1*/*GhTTG3* genes have overlapping functions with *GhTTG2*/*GhTTG4* in cotton. An example of common functionality between clades was revealed when comparing related bHLH proteins involved in anthocyanin biosynthesis of various plant species where (as is the case for WDR proteins) two major clades appear to exist (Esch *et al.*, 2003; Ramsay *et al.*, 2005; Morita *et al.*, 2006). One bHLH clade contains the *Arabidopsis* proteins GL3 and EGL3, whereas the second includes another *Arabidopsis* bHLH protein, TT8. Each of these proteins plays a common role in anthocyanin production, whilst GL3 and EGL3 have additional roles (trichome and root hair initiation) that differ from that of TT8 (seed coat mucilage production).

In a recent study of the flowering plant Japanese morning glory (*Ipomoea nil*), genes encoding MYB, bHLH and WD-repeat proteins involved in anthocyanin production were sought (Morita *et al.*, 2006). Two WD-repeat encoding genes were isolated, termed *InWDR1* and *InWDR2*. A phylogenetic analysis of these two protein sequences showed that InWDR1 was most closely related to AtTTG1 and AN11, whilst InWDR2 grouped with ATAN11. This may be similar to the situation in *G. hirsutum*, in which perhaps an additional duplication event has resulted in two genes from each clade being formed.

It appears that species contain either a pair of genes which fall into the PAC1-like and MP1-like groups, or into the ATAN11-like and TTG1-like groups. For example, for those species which contain a PAC1-like sequence (maize, rice and barley) there is also a sequence which groups closely with MP1. In addition, wheat and sorghum (which possess MP1-like sequences) do contain partial sequences with high homology to PAC1 (Carey *et al.*, 2004), although these are not included in this phylogenetic analysis. Similarly, partial ATAN11-like sequences exist for several species which are shown to contain only a TTG1-like sequence in Figure 3.5, such as tomato, poplar and iceplant.

Each of the species that contain PAC1-like and MP1-like sequences are grasses of the class Liliopsida, and belong to the Poaceae family. None of the grasses from this family appear to contain a WD-repeat sequence more closely related to TTG1 than to PAC1. It therefore appears that the PAC1 and MP1-related gene sequences are restricted to the Poaceae family of grasses, and are not present in any other plant family. It can thus be assumed that the PAC1 genes arose in a common ancestor of these species. The full-length sequences of PAC1-like proteins are of greater length than those of TTG1-like proteins (average length approximately 355 residues compared with 341). Similarly, MP1 proteins are on average around 50 amino acids longer than the ATAN1-like sub-group. Hence both MP1 and PAC1 appear to have evolved through the Poaceae family via insertions in ATAN11 and TTG1 respectively. There appears to be no species that contain both PAC1-like and TTG1-like sequences (or both MP1-like and ATAN11-like). These sub-divisions may therefore have occurred from the divergence of monocots and dicots, and the genes from each sub-group likely show redundant functions within their respective species. The fact that the *pac1* gene is able to complement multiple phenotypes in *Arabidopsis ttg1* mutants (Carey *et al.*, 2004) suggests the PAC1-like group of proteins has retained its functionality since this divergence.

In a study of the WD-repeat superfamily in *Arabidopsis*, van Nocker and Ludwig (2003) investigated the conservation of the structure and function of a range of WD-repeat proteins across different species. Proteins that were classified as homologous between species almost always showed the highest similarity within the WD-repeat regions. In some circumstances, the proteins showed little or no homology outside of the WD-repeat region. In this study by van Nocker and Ludwig, it was considered that even in the

examples of no homology outside the WD-repeat region, the function of the homologous protein would still be conserved. As observed in Figure 3.1, the highest homology between the GhTTG protein sequences and AtTTG1 is within the four WD-repeat regions. However, there is also considerable homology outside of the WD-repeats. Given the high homology observed both within and outside the WD-repeat regions between AtTTG1 and the GhTTG protein sequences, it is likely that function is conserved, based on the criteria of van Nocker and Ludwig. To test whether any of these four WD-repeat genes from cotton display conserved function with TTG1, cross-species complementation experiments were performed on *ttg1* mutants, and these are described in Chapter 4.

Chapter 4: Cross-species complementation of cotton WD-repeat genes

4.1 Introduction

The potential *TTG1* homologues identified in cotton displayed high sequence similarity to *Arabidopsis TTG1*. To test whether the observed sequence and structural similarities between the cotton WD-repeat proteins and *Arabidopsis TTG1* reflect functional similarity, the ability of the cotton genes to complement *ttg1* mutant phenotypes in heterologous systems was investigated.

4.2 Previous work

Previously (Humphries, 2001), a novel transient expression system was utilised to examine whether the cotton genes could complement a *ttg1* mutation in *Matthiola incana*. *Matthiola,* like *Arabidopsis*, is a member of the Brassicaceae, and a homologue of *TTG1* exists in *Matthiola* (*MiTTG1*) that affects both trichome and anthocyanin production (Kappert, 1949).

The anthocyanin biosynthetic pathway in the *ttg1* mutant *M. incana* is blocked at the dihydroflavanol-4-reductase (DFR) step (A. Walker, pers. com.), which results in a white-petal phenotype as compared to the purple petals of wild-type plants. The *ttg1* homologue within the *M. incana* mutant line 17 (*ttg1* mutant) has been sequenced, and shown to contain a lesion (substituting Arg with Trp) in one of the WD40 repeats (A. Walker, pers. com.). A second *Matthiola* mutant that has white petals is deficient for *chalcone synthase* (*CHS*) (Epping *et al.*, 1990), an enzyme downstream of *DFR* in the anthocyanin pathway, and this mutant was used as a negative control in the complementation assay described below.

The four cotton *TTG1*-like genes were cloned into the vector pART7 (Gleave, 1992; Appendix 1), such that their expression would be driven by the constitutive cauliflower mosaic virus (CaMV) 35S promoter. This promoter confers high-level transcription in all plant tissues (Harpster *et al.*, 1988), and hence enabled transient expression of the cotton *TTG1*-like genes after particle bombardment.

Particle bombardment was used to transiently express each of the four cotton genes and *Arabidopsis TTG1*, under the control of the 35S promoter, in petal tissues of the *M. incana ttg1* mutant. Bombardment with 35S::*AtTTG1* resulted in purple spots of anthocyanin production, indicating that *Arabidopsis TTG1* complements the mutation (Figure 4.1A). Purple spots of anthocyanin were observed by the naked eye approximately 40 hours after bombardment, and the anthocyanin spots spread and increased in intensity until approximately 60 hours after bombardment. Similar purple spots of anthocyanin production were visible on petals transiently expressing 35S::*GhTTG1* (Figure 4.1B) and 35S::*GhTTG3* (Figure 4.1C). None of the petals transformed with 35S::*GhTTG2* or 35S::*GhTTG4* showed any complementation, as assessed with the naked eye and after a series of careful searches with a dissection microscope at 40 and 60 hours after bombardment (Figure 4.1D,E). Bombardment of the *M. incana chs* mutant with *Arabidopsis TTG1* and any of the cotton *GhTTG* constructs did not produce purple spots (Figure 4.1F). The failure of *GhTTG1* and *GhTTG3* to complement a mutation in *chs* shows that the result is specific for *MiTTG1*.

This result suggests that *GhTTG1* and *GhTTG3* can function like *TTG1* in the production of anthocyanin pigments in *Matthiola*. However, it is the role of *AtTTG1* in trichome initiation which is of most interest in this study, and hence a stable transformation was performed in *Arabidopsis* to determine any function in trichome initiation.

4.3 Arabidopsis stable transformation

Stable transformation of *Arabidopsis ttg1-1* mutant plants was carried out via *Agrobacterium*-mediated transformation using constructs containing each of the four cotton WD-repeat genes.

4.3.1 Agrobacterium-mediated genetic transformation

The bacterial plant pathogen *Agrobacterium tumefaciens* provides a natural system to introduce genetic material into plant cells. It carries a large tumour-inducing (Ti) plasmid, a segment of which is designated T-DNA. This T-DNA segment is flanked by two 25 bp direct repeats and is transmitted by *A. tumefaciens* into individual plant cells as a single-stranded molecule in response to signals from the host plant (Thomashow *et al.*, 1980). These signals induce the expression of *Agrobacterium* virulence (vir) genes located on the

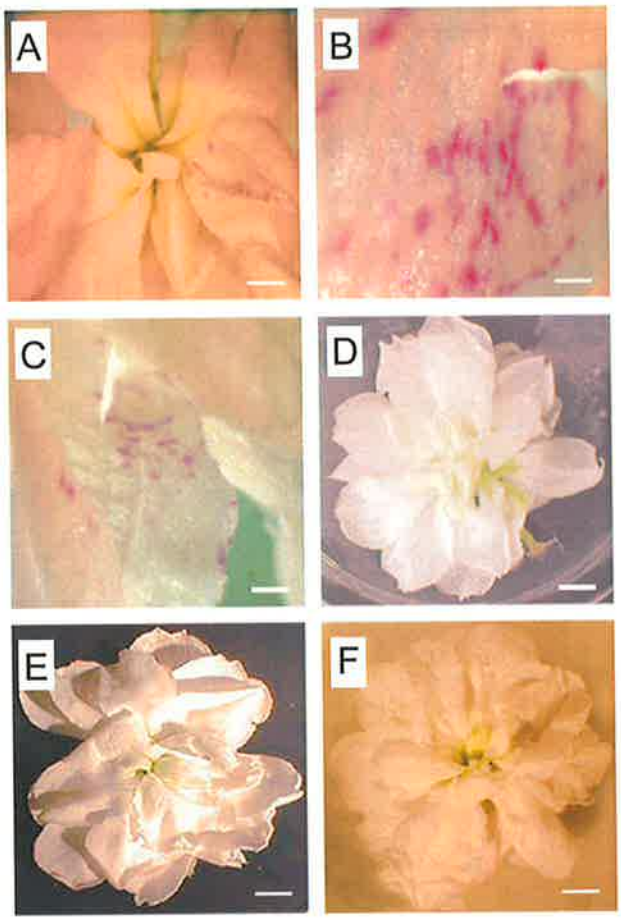


Figure 4.1: Particle bombardment of *Matthiola* flowers with *TTG1* gene constructs

(A) *M. incana ttg1* mutant petals bombarded with *Arabidopsis TTG1* positive control, showing purple spots of anthocyanin production.
(B) and **(C)** Patches of anthocyanin on *M. incana ttg1* petals transiently expressing cotton *TTG1*-like genes *GhTTG1* (B) and *GhTTG3* (C).
(D) and **(E)** *M. incana ttg1* mutant flowers bombarded with cotton *TTG1*-like genes *GhTTG2* (D) and *GhTTG4* (E) , showing no purple spots.
(F) *M. incana chs* mutant flower bombarded with *Arabidopsis TTG1*, showing no purple spots.
Bars = 3 mm in (A), 0.4 mm in (B), 1 mm in (C) and 5 mm in (D), (E) and (F)

Ti plasmid. A VirD1-VirD2 heterodimer then excises the T-DNA from the Ti plasmid, allowing export to the host cell. Once inside the plant cell, the T-DNA is targeted to the plant cell nucleus, where it is stably incorporated at random positions in the genome. In this way, *Agrobacterium* induces crown gall disease in many plant species, as genes inserted between the direct repeats function to divert the plant cell metabolism to provide for specific bacterial nutrition (Smith and Townsend, 1907). *Agrobacterium*-mediated genetic transformation (AMGT) takes advantage of this natural plant transformation system to stably transform plants with genes of interest. The T-DNA transfer is not sequence specific, so the native genes responsible for the formation of tumours can be replaced by any gene of interest (Leemans *et al.*, 1982). This results in the stable integration and inheritance of the desired gene within the plant genome after infection with genetically attenuated strains of *Agrobacterium*. In the past few years, there has been a sharp increase in the number of plant species reported to be transformed by *Agrobacterium*, facilitated principally by developments in tissue culture techniques specific for these plant species (reviewed in Lacroix *et al.*, 2006). AMGT has also been reported in a variety of non-plant cells, the first of which involved *Saccharoymces cerevisiae* (Bundock *et al.*, 1995), and followed by filamentous fungi and some human cell lines (de Groot *et al.*, 1998; Kunik *et al.*, 2001).

4.3.2 Transformation of Arabidopsis ttg1-1 mutant plants

For *Agrobacterium*-mediated delivery of a gene into plants, the DNA to be incorporated is commonly cloned into a binary vector to allow propagation in both *E. coli* and *A. tumefaciens*. In this study, the binary vector pART27, which contains the RK2 minimal replicon, was utilised (Schmidhauser and Helinski, 1985; Gleave, 1992). pART27 contains a kanamycin resistance gene, such that transformed progeny can be identified on antibiotic selection media.

Each of the four cotton WD-repeat genes and *Arabidopsis TTG1* was cloned into pART27. The expression cartridge of the pART7 construct, comprising the CaMV 35S promoter, the gene of interest, and the 3' transcriptional terminator region of the agrobacterial octopine synthase gene, was transferred as a *NotI* fragment into the binary vector. Once verified by sequence analysis, each of these constructs was electroporated (Walkerpeach and Velten, 1994) into *Agrobacterium* strain AGL-1 (Lazo *et al.*, 1991).

ttg1-1 mutant *Arabidopsis* plants were transformed with cotton WD-repeat genes and *Arabidopsis TTG1* by floral dip infiltration (Clough and Bent, 1998; see Materials and Methods). The *ttg1-1* allele displays several strong phenotypes which are caused by a premature stop codon located 25 amino acids from the C terminus of the protein (Walker *et al.*, 1999). Seeds from the *ttg1-1* mutants transformed with each of the cotton *TTG1*-like genes and the positive control were collected and plated onto kanamycin selection plates to select for transgenotypes.

Kanamycin-resistant progeny were observed for all genes except *GhTTG4*, for which no transformants were observed, despite several independent attempts, and the screening of over 16,000 seeds. Between 8,000 and 16,000 seeds were screened for each experiment, with successful transformation rates varying between 0.3% and 0.5%. Those seedlings which displayed resistance to kanamycin were transferred to soil and allowed to set seed. PCR was performed on genomic DNA prepared from putative transformants, with subsequent sequencing of the amplified product to verify the presence of the expected transgene. These primary transformants were monitored for trichome formation.

Transformants with the positive control *AtTTG1* demonstrated normal trichome formation as expected (Figure 4.2C) and plants transformed with the *GhTTG1* and *GhTTG3* genes also displayed normal trichome initiation and development (Figure 4.2D,F). *GhTTG2* transgenic plants remained entirely glabrous (Figure 4.2E). Trichomes of *GhTTG1* and *GhTTG3* transgenic plants demonstrated normal morphology, with the majority being three-branched (Figure 4.2G-L). Scanning electron micrograph (SEM) analysis of the trichomes formed on *GhTTG3* transgenic leaves demonstrated that regular trichome accessory cells had formed (Figure 4.2M-O), further indicating that normal trichome initiation had occurred in these plants.

Trichome density was assessed in several transgenic lines containing *35S::GhTTG1* and *35S::GhTTG3*. Counts were performed on the first four true leaves and compared with trichome production in mutants transformed with *35S::AtTTG1*. These counts showed that the 35S*::GhTTG1* and 35S*::GhTTG3* plants were similar and both had slightly decreased trichome numbers compared with those transformed with wild-type *AtTTG1* (Table 4.1). The number of initiating trichomes showed some level of variation between lines for a

given transgene, which may be attributed to positional effects associated with integration of T-DNA during AMGT.

Leaf Number:	1	2	3	4
35S::AtTTG1	7.2 ± 2.2	13.6 ± 2.3	18.8 ± 3.1	38.4 ± 5.5
35S::GhTTG1	5.4 ± 2.2	11.3 ± 2.1	17.2 ± 2.2	34.1 ± 4.9
35S::GhTTG3	5.3 ± 1.9	11.5 ± 1.8	17.0 ± 2.1	33.6 ± 4.4

Table 4.1: Trichome numbers on the first four true leaves (1-4) of *ttg1-1* mutants transformed with *AtTTG1*, *GhTTG1* and *GhTTG3*. Trichome numbers were derived by averaging the counts from plants of varying lines transformed with each transgene.

4.3.3 Additional phenotypes

Progeny of selfed *35S::GhTTG2* and *35S::GhTTG3 ttg1-1* transformants were further examined for complementation of the full range of mutant phenotypes.

4.3.3.1 Proanthocyanidin production in the seed coat

In the case of *GhTTG3*, seeds from all kanamycin-resistant transformants were brown in colour, similar to wild-type, indicating complementation of the *ttg1* proanthocyanidin defect in the maternally-derived seed coat (Figure 4.3D). By contrast, seed coat tannin production was not restored by the *GhTTG2* transgene (Figure 4.3C), with seeds displaying the yellow appearance associated with the transparent testa phenotype.

4.3.3.2 Seed coat morphology

Scanning electron micrograph (SEM) analysis was performed to analyse the seed coat morphology more closely. In the wild-type seed, differentiated cells in the outer epidermis contain mucilage and develop thick walls and central elevations known as columellae (Figure 4.4A). The *TTG1* gene is essential for columella development in the testa epidermis and *ttg1-1* mutants demonstrate a collapsed columella phenotype, which can be observed using SEM (Figure 4.4C). *GhTTG3* transgenic seeds demonstrated normal seed coat morphology and wild-type columella formation (Figure 4.4G). In contrast, a collapsed-columellae phenotype was observed in *GhTTG2* transgenic progeny (Figure 4.4E) resembling that seen in the *ttg1-1* mutant. The *GhTTG2* transgenic seeds displayed the general disorganisation of the *ttg1-1* mutant with regard to structure of the seed coat

(Figure 4.4D,F), whilst *GhTTG3* seeds were observed to have the more structured appearance characteristic of wild-type (Figure 4.4B,H).

4.3.3.3 Production of seed coat mucilage

Upon contact with water, the mucilage within wild-type seeds expands, rupturing the primary cell wall and extruding from the seed coat. The *TTG1* gene plays a role in mucilage extrusion during imbibition (Koorneef, 1981), and mucilage released from a seed can be detected by ruthenium red staining (Hanke and Northcote, 1975). Staining of imbibed seeds demonstrated that seeds transformed with *GhTTG3* produced releasable mucilage, characteristic of wild-type *Arabidopsis* (Figure 4.5A,D). No releasable mucilage was detected from *GhTTG2* transformed seeds, as observed for the *ttg1-1* mutant (Figure 4.5B,C).

4.3.3.4 Anthocyanin production in vegetative tissue

Wild-type *Arabidopsis* seedlings display anthocyanin accumulation in the hypocotyl and cotyledons (Figure 4.6A), whereas *ttg1* mutants lack anthocyanin (Figure 4.6B). Wild-type purple anthocyanins were observed in cotyledons and the hypocotyl of 5-day old *GhTTG3* transformants (Figure 4.6D), but not *GhTTG2* transformants (Figure 4.6C). When grown on sucrose-rich media, mature *GhTTG3* transformants display strong pigmentation in the leaves (Figure 4.6H), whilst the leaves of plants containing the *GhTTG2* gene lack visible anthocyanin (Figure 4.6G).

4.3.3.5 Regulation of root hair positioning

In wild-type *Arabidopsis* roots, only those cell files that are positioned above the junction of two underlying cortical cells produce root hairs (Dolan *et al.*, 1994, Figure 1.5). All root cell files in *ttg1-1* mutants are able to adopt the hair cell fate (Galway *et al.*, 1994), resulting in the production of root hairs in adjacent files (Figure 4.7B). *GhTTG2* transformant roots display this pattern; when viewed under the microscope, root hairs could be observed initiating in adjacent files (Figure 4.7C). In contrast, *GhTTG3* root hair positioning is wild-type, with each vertical hair cell file being separated by at least one non-hair cell file (Figure 4.7D).

SEM analysis confirmed complementation of the root-hair mutant phenotype in *GhTTG3* transformants (Figure 4.8C,D) and verified that the root-hairs of *GhTTG2* transformants

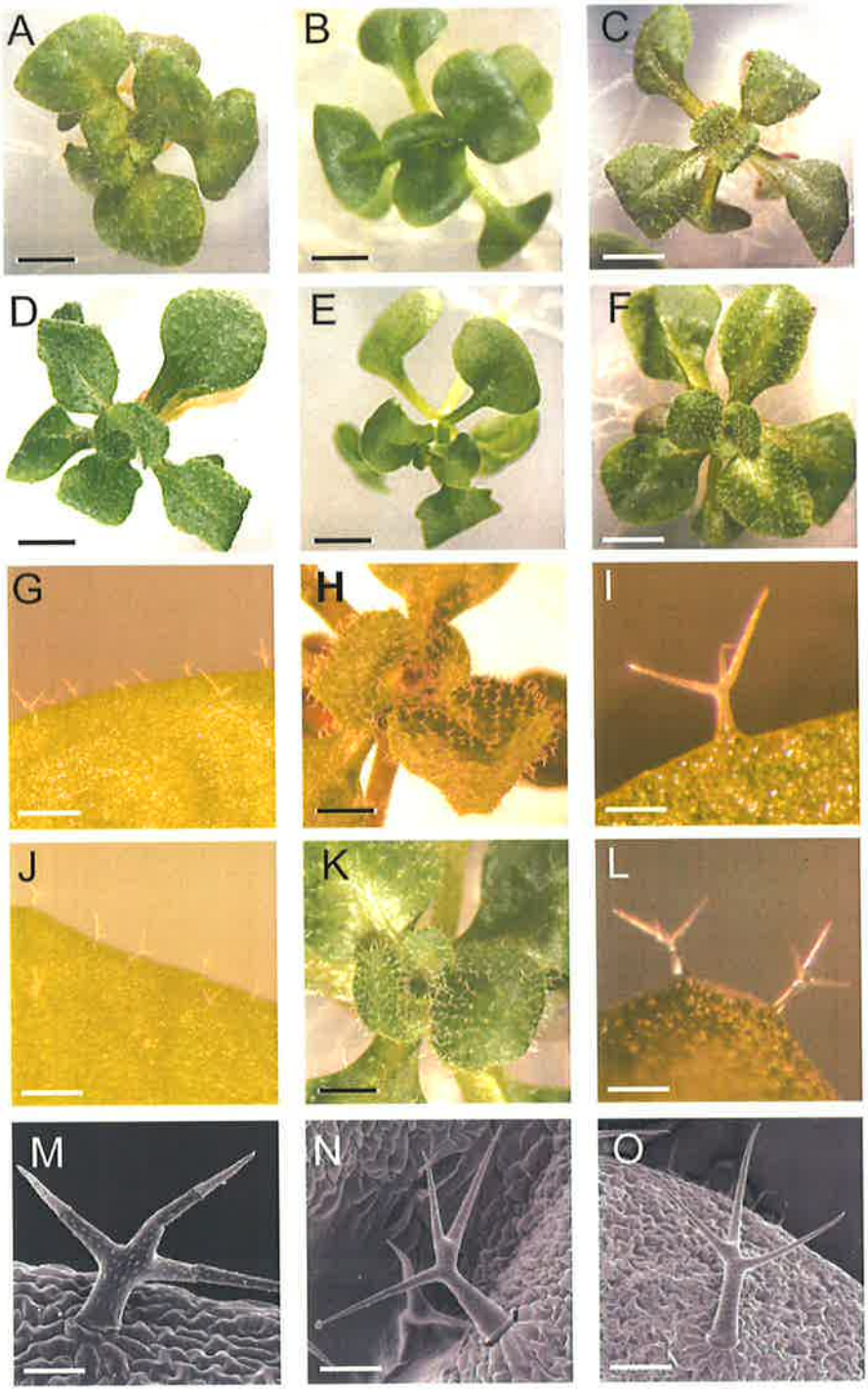


Figure 4.2: Phenotypes of *Arabidopsis* transformants and controls

(A-C) Wild-type (A), *ttg1-1* mutant (B) and *ttg1-1* transformed with *AtTTG1*.
(D-F) *ttg1-1* transformed with 35S::*GhTTG1* (D), 35S:*GhTTG2* (E), 35S::*GhTTG3* (F).
(G-L) Lines of 35S::*GhTTG1* (G-I) and 35S::*GhTTG3* (J-L) transgenic plants shown at various magnifications.
(M-O) SEM images of trichomes of 35S::*GhTTG3* transgenic plants.
Bar = 1 cm in (A), (C), (D), (E) and (F), 5 mm in (B), 500 μm in (G) and (J), 2.5 mm in (H) and (K), 100 μm in (I and O), 150 μm in (L), and 75 μm in (M) and (N).

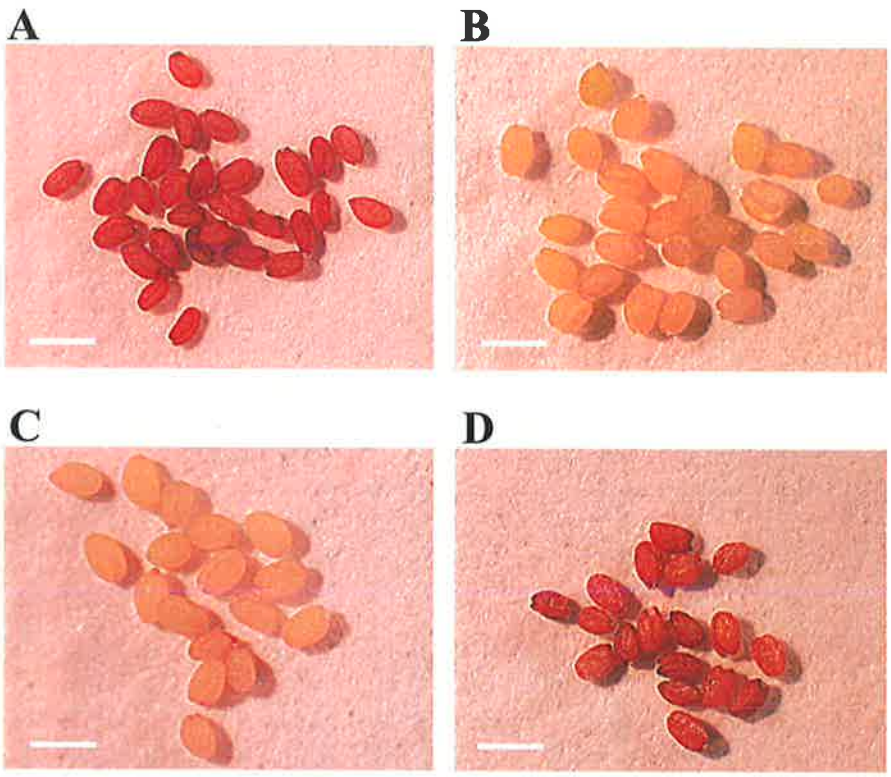


Figure 4.3: Seed coat pigmentation

(A) Wild-type *Arabidopsis* seed
(B) *ttg1-1* mutant seed
(C) 35S::*GhTTG2* transgenic seed, displaying the transparent testa phenotype
(D) 35S::*GhTTG3* transgenic seed, displaying wild-type proanthocyanidin production
Bar = 0.5 mm

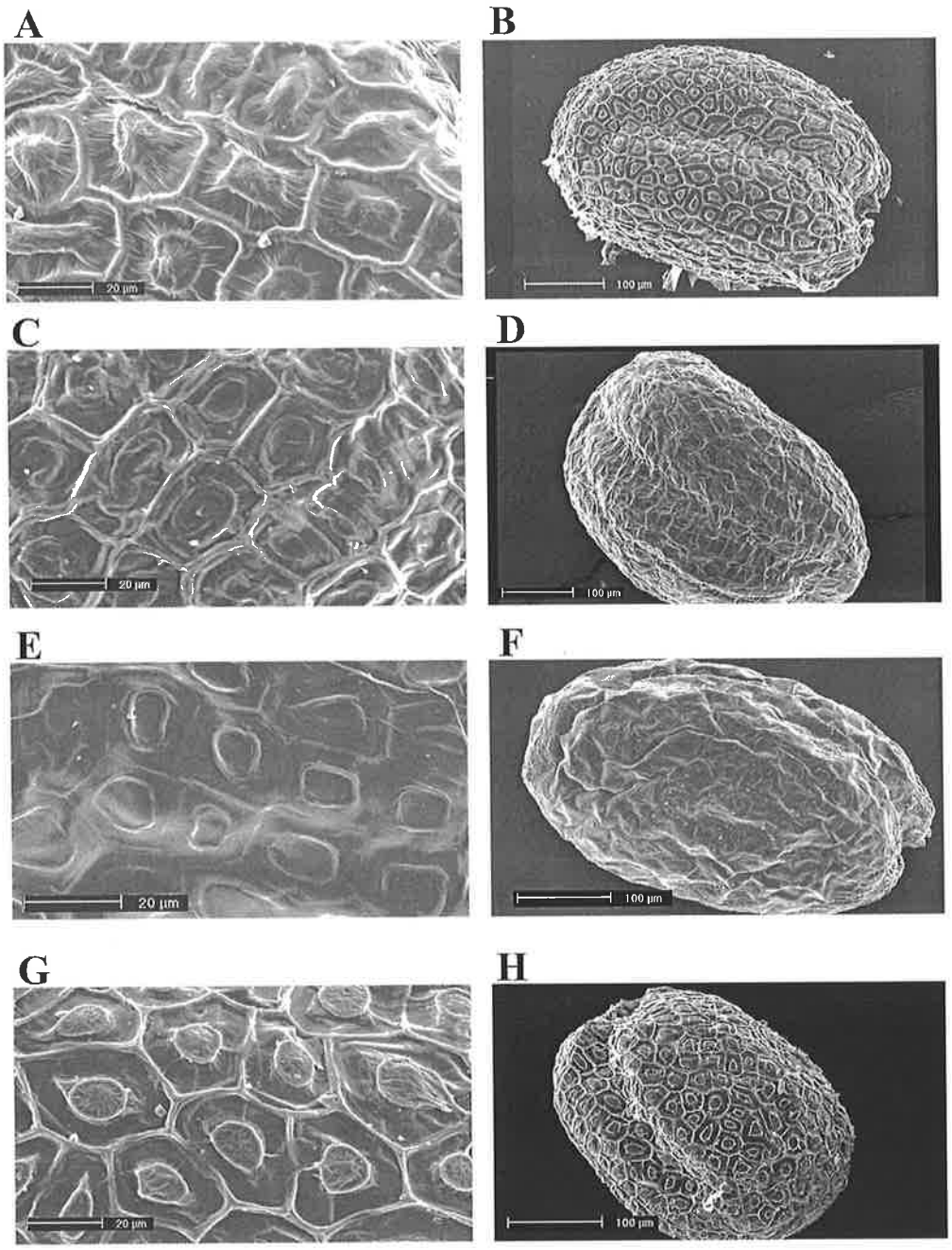


Figure 4.4: Seed coat morphology

SEM images of wild-type **(A, B)**, *ttg1-1* **(C, D)**, 35S::*GhTTG2* **(E, F)** and 35S::*GhTTG3* **(G, H)** seeds, demonstrating a magnified view of columella formation (Column 1), and overall seed coat structure (Column 2).

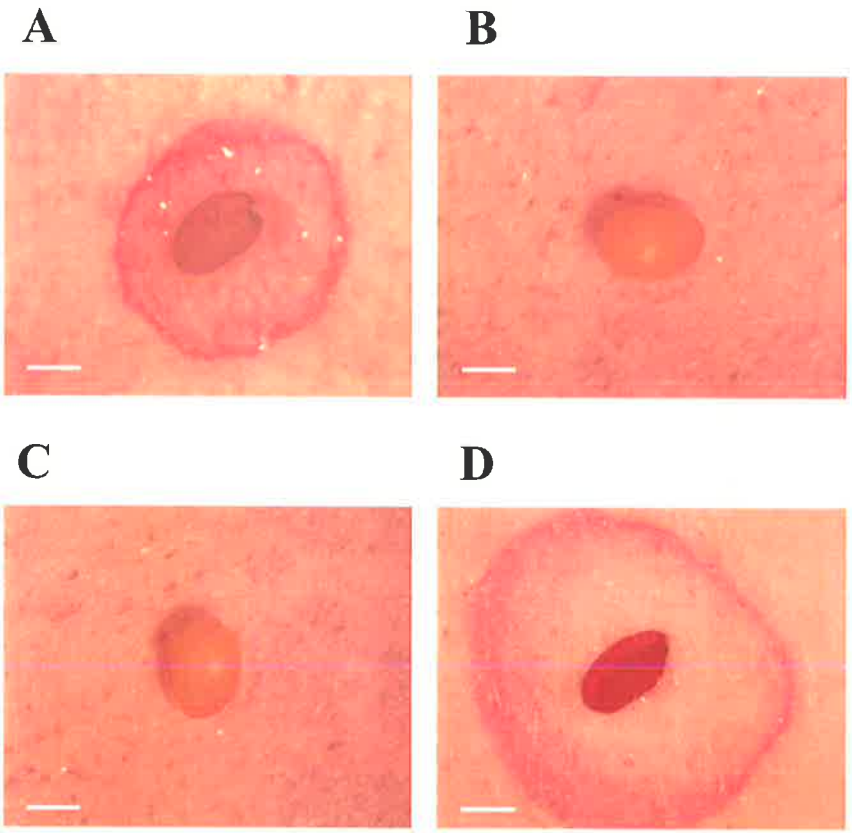


Figure 4.5: Releasable seed coat mucilage

Imbibed wild-type **(A)**, *ttg1-1* **(B)**, 35S::*GhTTG2* **(C)** and 35S::*GhTTG3* **(D)** seeds stained with ruthenium red; released mucilage is detectable by the presence of a red halo. Bar = 250 μm.

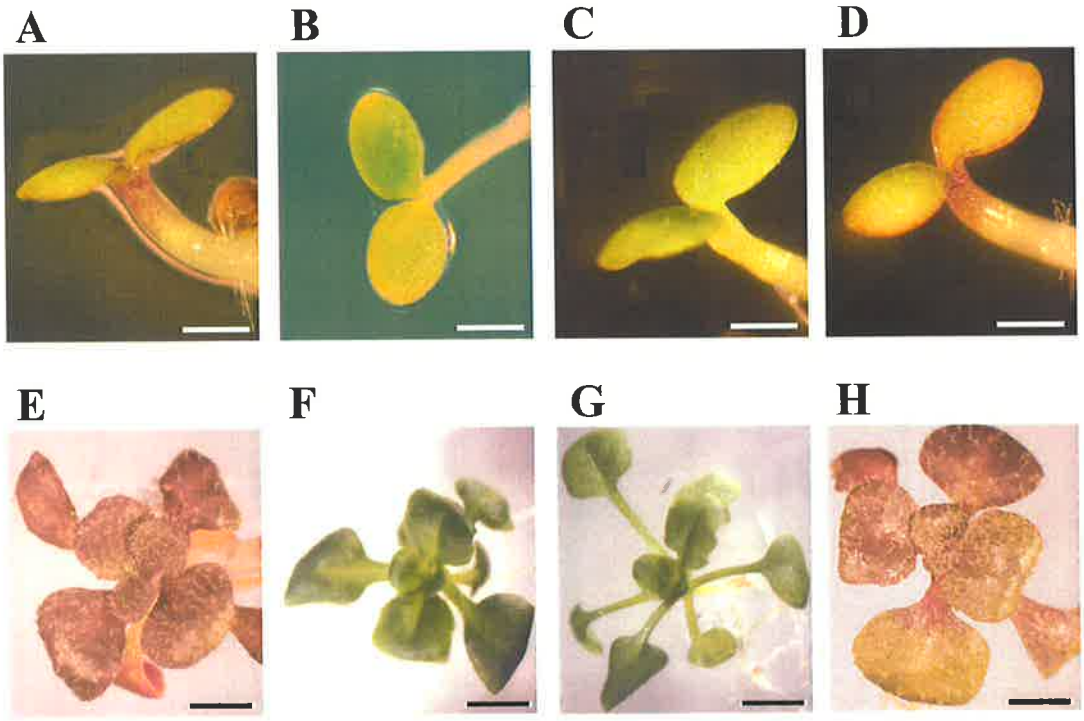


Figure 4.6: Anthocyanin production in leaves

(A-D) 5-day old wild-type (A), *ttg1-1* (B), 35S::*GhTTG2* (C) and 35S::*GhTTG3* (D) *Arabidopsis* seedlings, with anthocyanin accumulation apparent by purple colouration in the hypocotyl and cotyledons in images A and D.
(E-H) Mature plants grown on sucrose-rich media to induce anthocyanin production in wild-type (E), *ttg1-1* (F) *35S:GhTTG2* (G) and *35S:GhTTG3* (H) leaves.
Bar = 2 mm in (A-D) and 1 cm in (E-H)

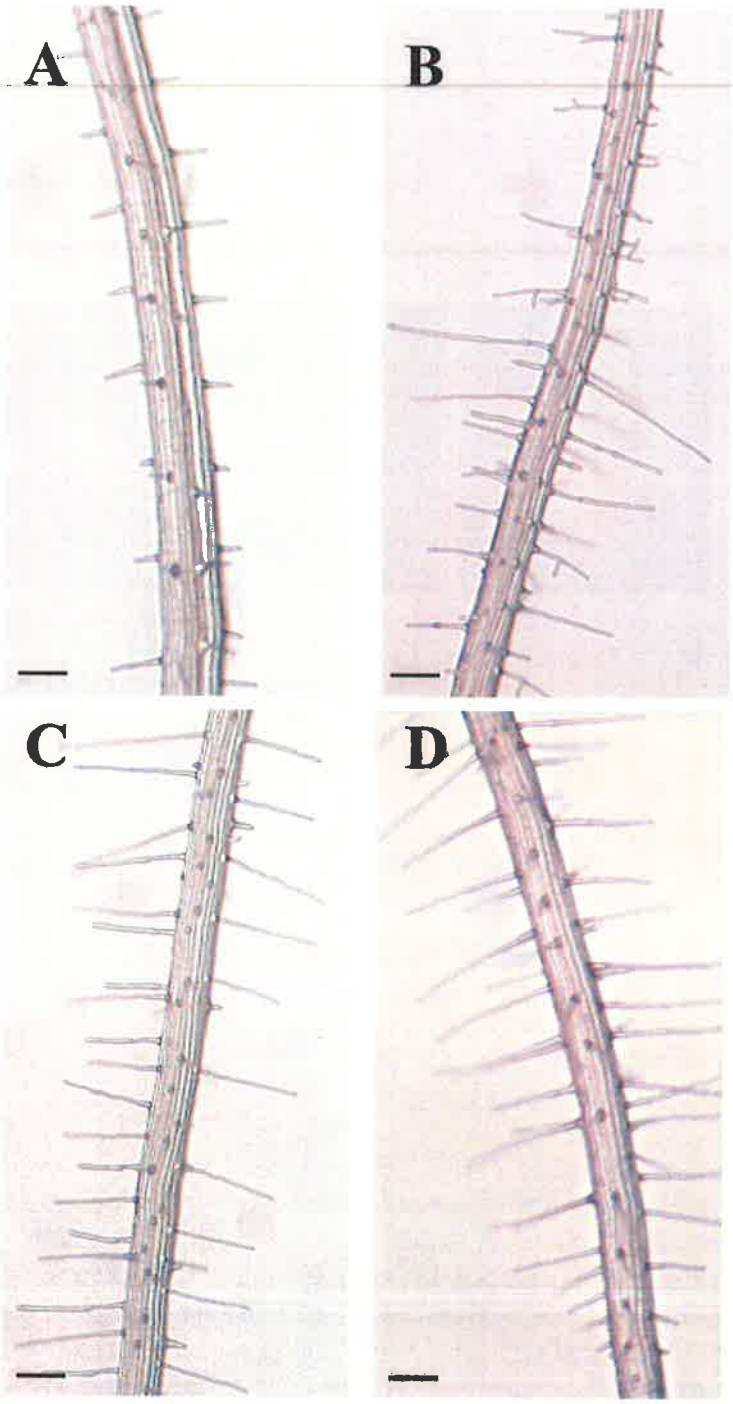


Figure 4.7: Root hair patterning

Light microscope images of wild-type **(A)**, *ttg1-1* **(B)**, 35S::*GhTTG2* **(C)** and 35S::*GhTTG3* **(D)** roots, with wild-type root hair spacing apparent in *35S:GhTTG3*, whilst root hairs initiate in adjacent vertical files in *35S:GhTTG2* (C), as observed in *ttg1-1*.
Bar = 100 μm in (A) and 150 μm in (B-D).

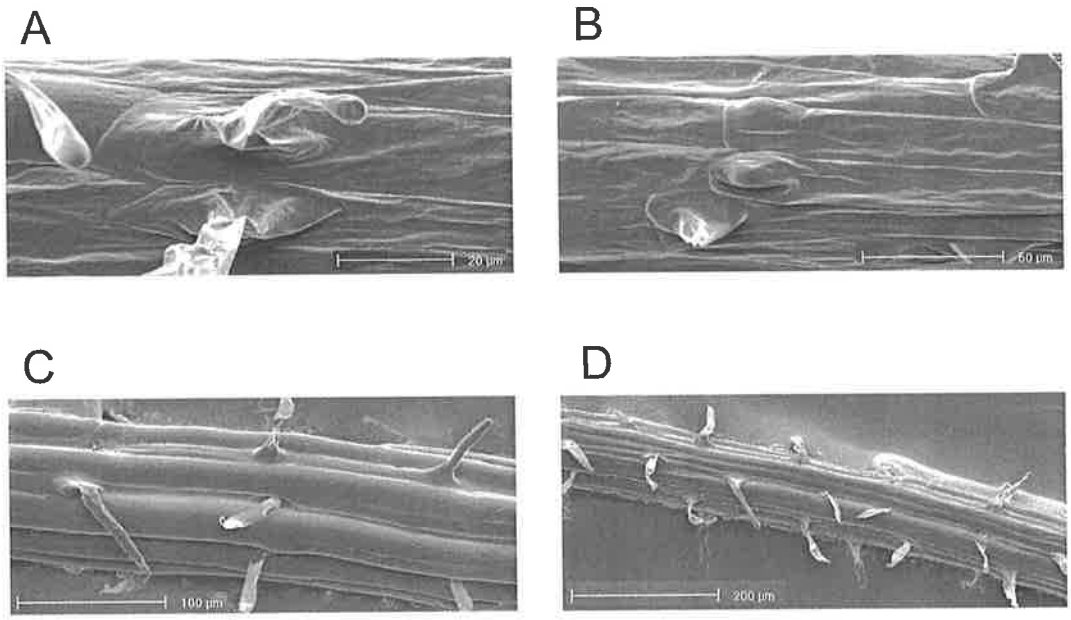


Figure 4.8: Root hair patterning

SEM images of 35S::*GhTTG2* **(A,B)** and 35S::*GhTTG3* roots **(C,D)**, with wild-type root hair spacing apparent in *35S:GhTTG3*, whilst root hairs initiate in adjacent cell files (shown horizontally in this figure) in *35S:GhTTG2*.

develop in adjacent cell files (Figure 4.8A,B). Due to their delicate nature, root-hairs collapse upon exposure to air, hence the root hairs in the SEM analysis appear distorted and in some cases have become separated from the root. However, the position of the root-hair base, demonstrating the pattern of root hair initiation in each transgenic line studied, can clearly be observed.

4.4 Discussion

Each of the four WD-repeat cotton genes were tested for functional homology to *TTG1*, by investigating complementation of the various mutant phenotypes in heterologous systems. In both the *M. incana* and *A. thaliana* investigations, two of the four genes, *GhTTG1* and *GhTTG3* were able to complement the mutant phenotype. These are the two genes with the highest sequence homology to *Arabidopsis TTG1*, and which fall into the same clade as *AtTTG1* and *MiTTG1* in the phylogenetic analysis of plant WD-repeat genes (Clade A, Figure 3.5), along with other genes known to complement *ttg1* mutations in *Arabidopsis*.

The transient assay in *M. incana* demonstrated that *GhTTG1* and *GhTTG3* were able to induce purple spots of anthocyanin in the *ttg1* mutant. Complementation was not confined to single cells, but groups of cells appeared to produce the purple pigment, allowing visualisation with the naked eye. It has been shown previously that complementation of a *M. incana* bHLH mutant (Ramsay *et al.*, 2003) by particle bombardment with either of two *Arabidopsis* bHLH genes (GL3 and MYC-146) results in only single cells containing visible anthocyanin. Similarly, Ludwig *et al.* (1990) demonstrated that the introduction of Lc (Leaf color) cDNA into tissues of maize by particle bombardment induces cell-autonomous pigmentation. Therefore, the multi-cellular spots observed on transient expression of *GhTTG1* and *GhTTG3* in *ttg1* mutants is unlikely to be due to clusters of cells being hit by gold particles, or diffusion of anthocyanin between cells, as this could have equally occurred in these previous investigations. The presence of multi-cellular spots in the *Matthiola ttg1* mutants in this experiment therefore suggests that cell-cell transfer of a signalling molecule has occurred. The TTG1 protein itself may move between cells, as a component of the putative diffusible trichome regulatory complexes (Zhang *et al.*, 2003) or it may regulate a biosynthetic intermediate that is transported between cells.

Of primary interest was the ability of the *GhTTG* genes to promote trichome formation in the *Arabidopsis ttg1-1* mutant background. A closer investigation of *GhTTG2* and *GhTTG3*

transgenic progeny demonstrated that *GhTTG3* was able to complement the full range of *ttg1-1* mutant phenotypes, whilst *GhTTG2* had no effect on any pathway examined. Although not as easily quantified as trichome number, these additional phenotypes appeared to be complemented essentially to wild-type phenotype in the *GhTTG3* transgenic lines. This has been the pattern observed in previous *ttg1* complementation experiments where, if one mutant phenotype is complemented, all phenotypes are complemented. For example, Carey *et al.* (2004) demonstrated that the maize *pac1* gene was able to complement each of the *ttg1* mutant phenotypes tested (although root hair patterning was not investigated). Similarly, the maize *R* gene was observed to complement all *ttg1* phenotypes (Lloyd *et al.*, 1992), as was the petunia gene *an11* (Payne, unpublished data). Prior to the cloning of the *TTG1* gene (Walker *et al.*, 1999), the molecular nature of the pleiotropy of *ttg1* mutants was unclear. One hypothesis postulated that *TTG1* encoded a protein with multiple independent functional domains, where mutations affecting each function might cluster in particular regions of the gene (Larkin *et al.*, 1999). However, an alternative hypothesis has proved to be correct, with *TTG1* encoding a protein which interacts with several different proteins, each of which controls a particular function (or set of functions). This became apparent as every allele of *ttg1* has the same pleiotropic phenotype, with all pathways affected to some extent. Additionally, all documented complementation experiments, including this study, demonstrate no segregation of phenotypes, with either none or all pathways being affected by an introduced transgene in *ttg1* mutants.

It is not known why positive transformants were unable to be recovered for *GhTTG4*. Whilst the sequence of the *GhTTG4* coding region in the binary vector was confirmed to be correct, an error may have occurred elsewhere in this construct that has interfered with transformation. As several independent transformations were attempted with this gene, the lack of transformants is unlikely to be due to any technical problem with the floral dip transformation method, particularly given the success observed with each of the other constructs. Although it is unknown whether *GhTTG4* can substitute for *Arabidopsis TTG1*, it appears unlikely given the high level (97%) amino acid similarity between protein sequences of GhTTG4 and GhTTG2 (which did not complement the mutant phenotype).

GhTTG2 and GhTTG4 demonstrate strong similarity to ATAN11A and ATAN11B, whose precise functions are not known, but are thought to play a role in anthocyanin production

(de Vetten, 1997). If this is the case, and ATAN11A or ATANB11B were involved in a TTG1-dependent pathway in anthocyanin production, it was considered possible that the *GhTTG2* transgene may have affected anthocyanin production in this investigation by influencing downstream targets in the pathway. This was not observed, with the most likely explanation being that ATAN11A and ATAN11B are not involved in a TTG1-dependent anthocyanin pathway. Hernandez *et al.* (2000) identified a gene similar to *GhTTG2* and *GhTTG4* in maize, in the search for a putative *An11/TTG1* homologue. This gene (*mp1*) also displays high sequence similarity to ATAN11A and ATAN11B and, as was observed in this study with *GhTTG2*, was unable to complement the mutant *ttg1* phenotype in *Arabidopsis*. As previously discussed, the *pac1* gene of maize was later shown to be more closely related to *TTG1*, and is able to complement *ttg1* mutant phenotypes (Carey *et al.*, 2004).

The ability of *GhTTG1* and *GhTTG3* to promote trichome initiation in a glabrous *Arabidopsis* mutant gives an indication that *GhTTG1* and *GhTTG3* may play a role in cotton fibre differentiation. However, it should be noted that heterologous systems might not reveal gene function reliably. In the particular case of TTG-related processes, there are many examples in which the regulatory genes have been shown to be interchangeable between different species. For example, the *TTG1* homologue from maize, *pac1*, was able to complement all mutant phenotypes tested when transformed into *ttg1 Arabidopsis* (Carey *et al.*, 2004), despite the fact that *pac1* mutants have no effect on maize trichome formation and that seed coat mucilage is not produced in maize. Similarly, whilst the effect of the *an11* mutation in petunia is restricted to anthocyanin production, *an11* is able to complement the *Arabidopsis ttg1* trichome mutant phenotype (Payne, unpublished data). Yeast and human WD-repeat genes homologous to *an11* partially complement the petunia pigmentation mutant in transient assays, even though the source species do not produce anthocyanins (de Vetten *et al.*, 1997). These results indicate a general functional similarity between members of this group of WD-repeat proteins, which does not always reflect the normal function of the endogenous protein.

There are clearly significant differences in the developmental regulation of trichomes of different species, and caution is required when extrapolating results from these systems to cotton fibres. Payne *et al.*, (1999) analysed the developmental similarities between cotton fibres and the trichomes of *Arabidopsis* and tobacco by over-expressing a number of cotton

MYB transcription factors in tobacco and *Arabidopsis*. Over-expression of *CotMYBA* produced an increased number of trichomes in tobacco, but had no effect on trichome production in *Arabidopsis*. Similarly, Wu *et al.*, (2006) isolated a cotton MYB gene which was down-regulated in a fibreless cotton mutant (*GhMyb25*), and displayed similarity to the snapdragon MYB gene *MIXTA* and, more distantly *AtGL1*. This gene was over-expressed in *Arabidopsis* and tobacco, and its effects on trichome formation investigated. The expression of *GhMyb25* in *Arabidopsis* had no obvious effect on trichome development. Tobacco lines transformed with *GhMyb25* demonstrated increased numbers of branched long-stalked trichomes that are not usually observed on the adaxial leaf surface, another sign of the differing genetic control of trichome formation between different plant species (see also Serna and Martin, 2006). Moreover, none of the fibreless cotton mutants which have been studied show alteration in leaf trichome development, indicating that different genetic machinery controls cotton leaf trichomes and fibres. These findings suggest that some similarity exists in the molecular control of cotton fibre and leaf trichome initiation, but there is not a direct conservation between the two systems.

The use of *Arabidopsis* trichomes as a model system of cotton fibre development has prompted other investigators to perform complementation experiments similar to those described in this chapter. For example, Wang *et al.* (2004) performed an investigation of genes involved in fibre initiation, searching for a cotton homologue of the *AtGL1* gene. Of the *MYB* genes isolated in this study, the most interesting candidate, *GaMYB2*, was tested for functional complementation of glabrous *Arabidopsis gl1* mutant. It was demonstrated that this cotton gene, which shares approximately 60% sequence identity to *AtGL1*, could rescue trichome formation in the *gl1* mutant, under the control of the *GL1* promoter. Under the control of the 35S promoter, *GaMYB2* was even able to induce trichomes (about 1-4 per seed) on *Arabidopsis* seeds (which are usually glabrous) in both the wild-type and *gl1* backgrounds. The presence of trichomes on the transgenic *Arabidopsis* seed appears to be a good indicator of a role of this particular gene in cotton fibre initiation, as trichomes originating from the seed are likely to be very similar structures to cotton fibres, given that fibres also arise from the ovule epidermis. No seed trichomes were observed in this investigation.

4.5 Summary

Each of the four cotton *TTG1*-like genes was tested for their ability to replace the function of the *TTG1* gene in heterologous systems. In both *Matthiola* and *Arabidopsis*, the two cotton genes with the highest similarity to *AtTTG1* (*GhTTG1* and *GhTTG3*) were able to complement the *ttg1* mutant phenotype. One of these genes, *GhTTG3*, was shown to complement the full range of *ttg1* mutant phenotypes. These results give an indication that the protein products of these cotton WD-repeat genes act in a similar manner to the AtTTG1 protein, and may play a role in cotton fibre initiation. The fact that the two cotton genes could promote trichome initiation in *Arabidopsis* also provides further validation of the use of *Arabidopsis* trichomes as a model system for the study of cotton fibre development.

Chapter 5: Temporal and spatial expression of WD-repeat genes in cotton

5.1 Introduction

The functional analysis of a gene requires a detailed characterisation of its expression pattern at the tissue and cellular level. To characterise further the putative *TTG1* homologues in cotton, their transcription levels in various tissues were investigated. Genes found to be expressed in the epidermal cells of the cotton ovule, and specifically in fibre cells, are likely to have some role in fibre development.

5.2 Reverse-transcriptase PCR analysis

Reverse transcriptase PCR (RT-PCR) was performed to obtain an assessment of mRNA accumulation following transcription of the four WD-repeat genes in various tissues of the cotton plant. Previously attempted Northern blot analysis produced only weak or undetectable signals, presumably due to low mRNA levels of the WD-repeat genes (data not shown). Therefore, the more sensitive technique of RT-PCR was employed.

The gene-specific primers described in Section 3.2.1 were used to test for mRNA expression of each of the four *TTG1*-like cotton genes. Using these primers, RT-PCR was carried out on first-strand cDNA obtained by reverse transcription of total RNA derived from cotton fibres (at 6, 12, 18 and 24 DPA), ovule, leaf, flower, stem and root. Primers designed to amplify a 124 bp fragment of actin mRNA (Shimizu *et al.*, 1997) were used as an internal control to indicate the amount of cDNA in each sample. In addition, samples generated without the use of RT were tested to ensure that PCR products were not amplified from contaminating genomic DNA in the RNA samples.

As shown in Figure 5.1, each of the four *GhTTG* genes displayed similar mRNA accumulation profiles. In each assay, expression was observed throughout fibre development, and in all other tissues tested. No amplification was observed from non-reverse transcribed controls, or in the negative control with no template. Each of the genes was expressed in whole ovules on the day of anthesis as well as throughout the expansion and elongation stages of fibre development, and transcripts persisted into secondary cell

wall synthesis. The band produced in each of the four reactions from leaf mRNA as template were of lower intensity than the other tissues tested, but this was also the case in the actin control, indicating that *GhTTG* transcription in the leaf is most likely comparable to the other tissues. RT-PCR can provide only an indication of relative levels of expression and is not a quantitative measure. However, it appears that each of the *GhTTG* genes is expressed at an approximately equal level throughout fibre development, and in all other tissues tested.

5.3 Cellular localisation of cotton WD-repeat genes within the ovule

Whilst the RT-PCR analysis revealed no tissue-specific expression pattern of the *GhTTG* genes throughout the cotton plant, expression within the cotton ovule was further investigated. The cotton ovule is a complex structure, consisting of the nucellus, inner integument and the outer integument, from which the fibres are derived. Expression in fibre initials is an important prerequisite of a gene associated with fibre differentiation, and *in situ* hybridisation analysis was performed to determine whether the *GhTTG1-4* transcripts are present in specific cell types or generally throughout the cotton ovule.

5.3.1 Technical aspects of in situ hybridisation

In situ hybridisation is a powerful molecular method, which has been increasingly applied to locate mRNAs and characterise expression at the cellular level (Dagerlind *et al.*, 1992). *In situ* hybridisation can demonstrate the presence of a particular RNA or DNA sequence, and it differs from RT-PCR in facilitating the detection and localisation of specific nucleic acid sequences directly within a specific cell type (Gall and Pardue, 1969; John *et al.*, 1969).

In situ hybridisation experiments can be performed using either DNA or RNA probes that are complementary to the mRNA of interest, which can be chemically or radioactively labelled. RNA probes are advantageous since they form stable hybrids (more stable than DNA-RNA hybrids), are single-stranded, have little or no re-annealing during hybridisation, are highly specific, and display reduced background signals (Cox, 1968).

In recent times, the use of non-radioactive labelling has been preferred due to the ease and safety of handling, greater stability, and the shorter exposure times required (Farquharson *et al.*, 1990; Morris *et al.*, 1990). Another advantage of non-radioactive detection is that the

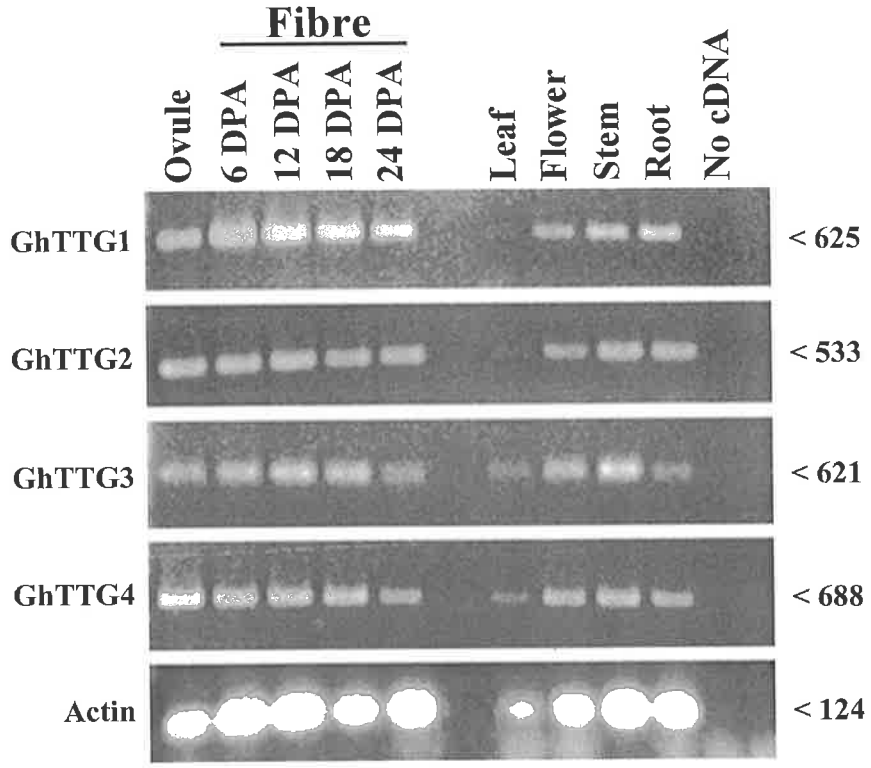


Figure 5.1: RT-PCR analysis of expression patterns of *TTG1*-like genes in cotton

Each lane contains equivalent loadings of PCR products obtained after initial reverse transcription of cotton tissue RNA, as indicated. Numbers on the right indicate the size of the PCR product in bp. Expression is observed for each of the four cotton genes throughout all organs tested.

cells expressing the gene of interest can be more precisely identified, allowing greater resolution than is possible with radioactive detection methods. The most commonly used non-radioactive detection methods utilise biotin and digoxigenin (DIG) labelling (Morris *et al.*, 1990). A disadvantage encountered in the use of biotin-labelled probes is the fact that many tissues contain endogenous biotin, thus leading to false positive results (Varma *et al.*, 1994). DIG-labelled probes are more sensitive (Chevalier *et al.*, 1997), and their use avoids the problems associated with endogenous biotin, as digoxigenin is only synthesised in plants of the genus *Digitalis* (Farquharson *et al.*, 1990). The DIG system is a widely used and effective method for the labelling of DNA, RNA and oligonucleotides, and it was utilised to generate labelled RNA probes in this study.

In preparation for the DIG-labelling process, the DNA used to generate the probe is required to undergo an *in vitro* transcription reaction. The DNA to be transcribed is cloned into the polylinker site of an appropriate transcription vector containing RNA polymerase binding regions (typically T7 and SP6 promoters). This construct is then linearised at a suitable restriction enzyme site downstream of the cloned insert, with preference given to enzymes leaving a 5' overhang. The RNA polymerases are then used to produce run-off transcripts, into which DIG-UTP is incorporated. Typically, one in every twenty nucleotides of the newly synthesised RNA is a DIG-UTP. DIG-labelled probes hybridise with nucleic acids *in situ* and can be detected with an anti-digoxigenin antibody conjugated to alkaline phosphatase, which is subsequently visualised through the NBT/BCIP colorimetric reaction. BCIP (5-bromo-4-chloro-3-ondolyl phosphate) is a substrate of alkaline phosphatase which gives a dark-blue indigo dye as an oxidation product (Horwitz *et al.*, 1966). NBT (Nitro-Blue Tetrazolium Chloride) serves as the oxidant and also gives a blue dye such that those cells in which the target gene is expressed can be identified by dark blue colouration (Wolf *et al.*, 1973).

5.3.2 Generation of in situ hybridisation RNA probes

Probes were designed to target the subgroups of *GhTTG1/GhTTG3* and *GhTTG2/GhTTG4*, as the similarity within each of these subgroups was too high to be able to generate specific individual probes. A representative gene from each subgroup was selected from which the probe was generated. The full-length sequences of *GhTTG2* and *GhTTG3* were cloned into pGEM-T Easy, which is a suitable vector for the subsequent *in vitro* transcription reaction as it contains the T7 and SP6 RNA promoter regions flanking the insertion site for PCR

products. These constructs were linearised by restriction enzyme digestion downstream of the insert. The restriction enzymes used were chosen on the basis that they create a 5' overhang and no other recognition sites were present in the construct (Table 5.1).

Probe	Construct	Enzyme used in linearisation	RNA polymerase utilised
GhTTG2 Anti-sense	TTG2IS4	*Spe*I	T7
GhTTG2 Sense	TTG2IS4	*Nco*I	SP6
GhTTG3 Anti-sense	TTG3IS2	*Spe*I	T7
GhTTG3 Sense	TTG3IS4	*Spe*I	T7
rRNA Anti-sense	rRNAS8	*SpeI*	T7
rRNA Sense	rRNAS1	*SpeI*	T7

Table 5.1: Details of each *in vitro* transcription reaction for the generation of the sense and anti-sense probes for *in situ* hybridisation analysis of cotton ovules.

The anti-sense RNA probe which is complementary to the target mRNA sequence is used to detect gene expression. A sense strand negative control is performed in parallel, to give a background signal approximating the level of non-specific interactions.

For technical reasons, the method used to generate sense and anti-sense probes was slightly varied between *GhTTG2* and *GhTTG3*. For *GhTTG2*, the same construct (termed TTG2IS4) was used for the production of both sense and anti-sense probes, with a different RNA polymerase and 3' restriction enzyme being used for each *in vitro* transcription reaction. In the case of *GhTTG3*, two different constructs were generated (TTG3IS2 and 4), containing the cloned insert in opposite orientations. The T7 RNA polymerase was used in both *in vitro* transcription reactions of these constructs to create a sense and antisense probe. In this case, the same 3' restriction enzyme (*Spe*I) could be used to linearise both plasmids (Table 5.1). A probe designed to the cotton 18S ribosomal RNA gene, which is known to be expressed at high levels throughout the ovule, was used as a positive control. A segment of the cotton *rRNA* gene was cloned into pGEM-T Easy in opposite orientations and both of these constructs were digested with *Spe*I, and the T7 RNA polymerase was used in the *in vitro* transcription reactions to create rRNA sense and anti-sense probes (Table 5.1). After DNase treatment, carbonate hydrolysis was performed on each of the probes to generate probe lengths of approximately 150 bp, because RNA molecules of this

length are ideal for tissue penetration and bind to target sequences faster than long molecules.

As *in situ* hybridisation using the DIG-labelling system is a very sensitive technique, it is important to work with defined amounts of probe and template to ensure optimal results. It is also critical to check the efficiency of each labelling reaction to ensure that the correct amount of probe is added to the hybridisation solution. This can be achieved by performing a spot test, which compares a standard of known concentration to a generated probe of unknown concentration. 0.5 µl of each labelled probe was spotted at varying dilutions onto a nylon membrane, and compared with the control (100 ng/µl) followed by colorimetric detection using NBT/BCIP. Each of the labelled probes was comparable to the supplied control labelled RNA, demonstrating that the DIG-labelling process had been effective, and allowed for estimates to be made of probe concentration enabling the appropriate amount of probe to be used in the hybridisation experiment.

5.3.3 In situ hybridisation results

In situ hybridisation was performed on paraffin-embedded sections of cotton ovules, which were prepared as described in Houben *et al.* (2005) (see also Materials and Methods). The *rRNA* sense and anti-sense probes were applied separately to 0 DPA sectioned ovules. The sense *rRNA* probe was also applied as a negative control. Probing with the *rRNA* antisense probe resulted in a strong purple staining throughout all cells after only 10 minutes of incubation in the NBT/BCIP colorimetric reaction, indicating the probe preparation and *in situ* hybridisation protocol was successful for this highly abundant RNA (Figure 5.2A,B). The sense control of the rRNA probe displayed no staining, even after an overnight incubation (Figure 5.2C).

Cellular expression of the *GhTTG2/GhTTG4* sub-group in 0 DPA and 2 DPA ovules was analysed using the *GhTTG2* probes. The method used for the ribosomal RNA positive control was replicated using the *GhTTG2* probes. However, neither the sense nor antisense *GhTTG2* probes detected any transcripts even after an overnight incubation (Figure 5.2D-F).

Expression of the *GhTTG1/GhTTG3* genes was investigated by hybridising 0 and 2 DPA sectioned ovules with the *GhTTG3* anti-sense probe. Hybridisation with the *GhTTG3* anti-sense probe produced a signal after approximately one hour of incubation in the NBT/BCIP colorimetric assay (Figure 5.2G,H,J,K). Probing with the *GhTTG3* sense probe produced no signal (Figure 5.2I,L), indicating that the staining observed using the *GhTTG3* anti-sense probe was specific for the *GhTTG3*-like mRNA.

The dark-blue staining observed for the *GhTTG3* anti-sense probe appeared to be restricted to the epidermal layer of cells and the cells with the darkest staining appeared to be those destined to become fibre cells. At 0 DPA, it was observed that the fibre initials that are enlarging, and beginning to protrude from the surface, displayed the strongest signal, exhibiting a very dark blue stain. A weaker signal was observed in the epidermal cells that do not differentiate into fibre cells, and in other cells of the outer integument. At 2 DPA, the fibre cells display expression of *GhTTG3* mRNA. This is consistent with a role for *GhTTG1/GhTTG3* in fibre development, specifically in fibre initiation.

Several explanations may exist for the fact that no expression was observed for *GhTTG2/4*, whilst signals could readily be observed using the *GhTTG1/3* anti-sense probe. Firstly, it may simply be that the *GhTTG2/4* mRNAs are less abundant, making them undetectable via *in situ* hybridisation. Another explanation may be that the *GhTTG2/4* genes are expressed in a more diffuse pattern in the ovule than *GhTTG1/3*, thus their expression is dispersed throughout the ovule at a level undetectable by *in situ* hybridisation, whereas for *GhTTG1/3*, the localisation of the mRNA to a specific cell type increases mRNA concentration in that cell type leading to a readily detectable level.

The *rRNA* positive control exhibited staining throughout the ovule, and no particular cell type within the ovule appeared especially disposed to staining during this procedure. Additionally, the sense control of *GhTTG3* demonstrated no staining in fibre cells or any cell type. These results suggest that *GhTTG1*, *GhTTG3* or both genes are expressed predominantly in initiating fibre cells, providing support for their putative role in promoting fibre growth.

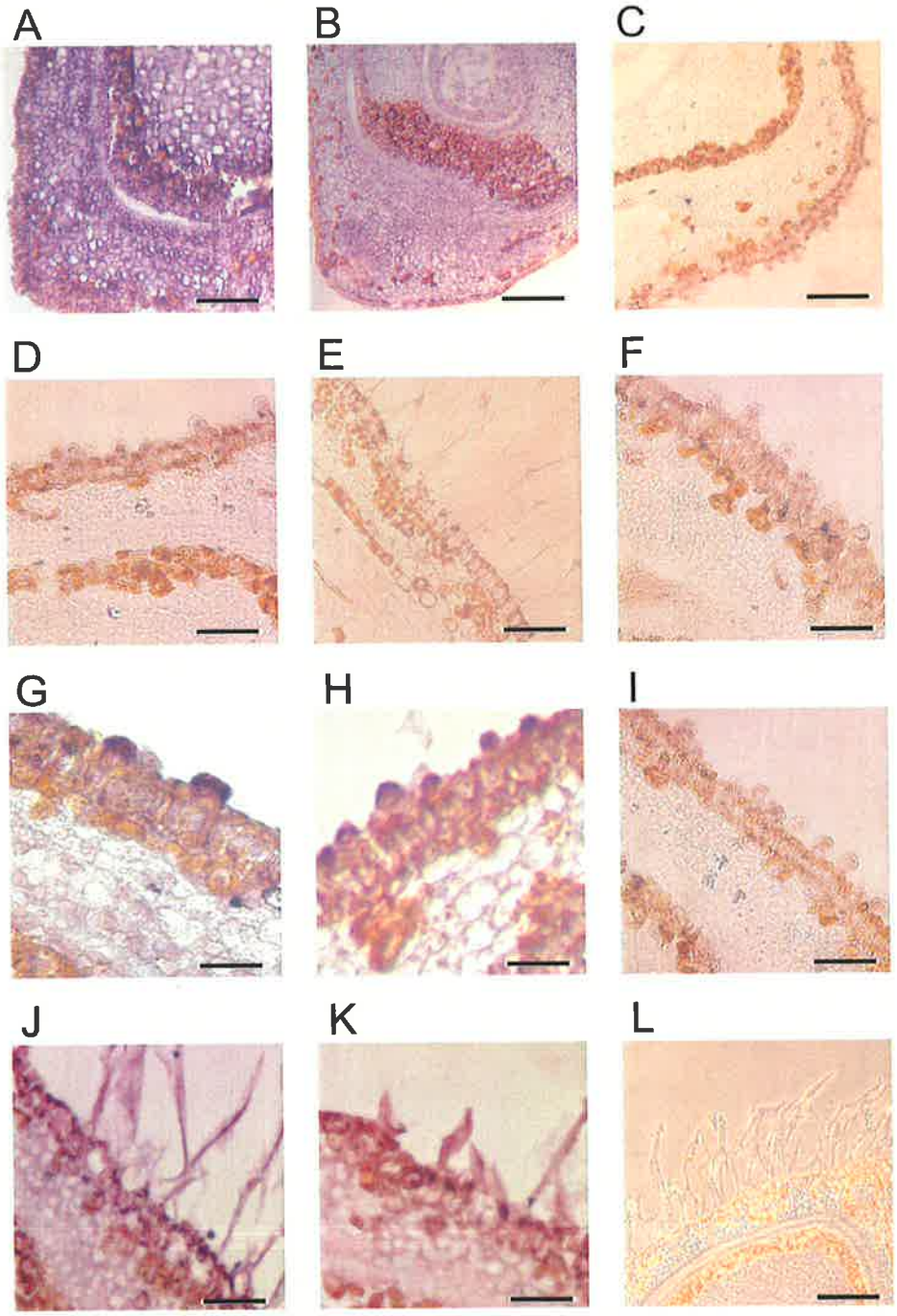


Figure 5.2: *In situ* hybridisation of sectioned cotton ovules
Purple staining indicates mRNA expression.
(A-C) Cross-sections of 0 DPA ovules hybridised with the *rRNA* anti-sense (A & B) and sense (C) probe.
(D-F) *GhTTG2/4* anti-sense probe hybridised to 0 DPA (D) and 2 DPA (E) ovule, and sense negative control (F). In each case, no purple staining was observed.
(G-I) 0 DPA ovules probed with *GhTTG1/3* anti-sense (G-H), and sense control (I). When anti-sense probe is applied, fibre initials display the darkest purple signal. This signal is absent in the sense control.
(J-L) 2 DPA ovules probed with *GhTTG1/3* anti-sense (J-K), and sense control (L). Purple staining can be observed in the epidermal and developing fibre cells of sections hybridised with the anti-sense probe, with little to no signal observed in other regions of the ovule. No signal was observed when applying the sense probe.
Bar = 100 μm in (A) and (L), 200 μm in (B) and (C), 40 μm in (D), (E), (F), (H), (I), (J) and (K) and 15 μm in (G)

5.4 Analysis of GhTTG expression in fibreless cotton mutants

To analyse further the expression of the cotton genes in the cotton ovule and to confirm the results observed in the *in situ* hybridisation experiments, an *in situ* analysis of fibreless mutants was performed.

5.4.1 Fibreless mutants of cotton

Numerous naturally occurring lintless mutants have been discovered. Some mutants have been found with normal lint, but absence of fuzz, whilst lintless mutants always lack fuzz (Turley and Kloth, 2002). The fuzzless seed phenotype has obvious advantages, due to ease of ginning. However, the fuzzless lines identified have a lower lint yield and so are not used in agriculture. Efforts are being made to identify the alleles responsible for lintless and fuzzless phenotypes (Turley and Kloth, 2002).

Lintless mutants are an excellent resource to analyse genes which may function in fibre development. One of the cotton fibreless mutants more widely used in genetic analysis is the *G. hirsutum* cv. Xu-142 *fuzzless-lintless (fl)* line (Zhang and Pan, 1991). This line is a naturally occurring mutant, originally discovered in a cotton field in China. SEM analysis has demonstrated that ovules of Xu-142 *fl* show virtually no fibre initiation (Yu *et al.*, 2000b), making it an excellent system for comparative analysis, together with wild-type Xu-142.

Numerous researchers have attempted to isolate and analyse genes differentially expressed between wild-type cotton and fibreless mutants. Ruan and Chourey (1998) characterised the cotton sucrose synthase gene (*Sus*) and showed via *in situ* hybridisation that its expression is reduced in the ovular epidermal cells of a fibreless mutant, providing evidence of a role in fibre development. Ji *et al.* (2003) used subtractive PCR and cDNA macroarray to identify genes involved in fibre elongation, comparing wild-type fibre cDNA with cDNA isolated from a Xu-142 *fl* mutant. Similarly, Li *et al.* (2002a) compared the expression profiles of ovules from Xu-142 and its *fl* mutant and identified genes that are preferentially expressed in elongating fibres. A recent investigation into cotton fibre initiation genes compared the transcription profiles of wild-type 0 DPA whole ovules with those of six fibreless (or fibre-reduced) lines using cDNA microarrays, unveiling several cases of differential transcription (Wu *et al.*, 2006). The identification of putative transcription factors and phytohormonal regulators acting during the early stages of fibre

development was facilitated by a comparison of transcript abundance with a fibreless mutant, *N1N1* (Yang *et al.*, 2006). These studies demonstrate the value of fibreless mutants in both the characterisation and identification of genes with a potential role in cotton fibre development.

5.4.2 Genomic PCRs on fibreless mutants

The first step of this analysis was to detect the presence of the *GhTTG* genes in a number of the fibreless mutants. Each gene was amplified from genomic DNA of the range of fibreless mutants available (see Materials and Methods). In each case, the primers amplified a product of expected length, indicating that no major alterations had occurred in the cotton *GhTTG* genes in the fibreless mutants (results not shown). Sequence analysis of a selection of the amplified PCR products demonstrated no sequence variations in any of the mutants, suggesting that none of the *GhTTG* genes were responsible for the fibreless phenotypes observed. The next step was to measure mRNA levels of the *GhTTG* genes in the fibreless mutant, to investigate any differential expression between wild-type and fibreless samples.

5.4.3 Reverse transcriptase-PCR analysis of a fibreless mutant

RT-PCR analysis was performed on wild-type Xu-142 and its fibreless mutant, to determine if the lack of fibre growth in the mutant has an effect on *GhTTG* mRNA expression. To perform this assay, ovules at the stage of flowering (0 DPA) and 3 DPA were dissected from wild-type and fibreless cotton bolls and RNA was prepared. First strand cDNA synthesis was performed on each sample by reverse transcription, using the Moloney Murine Leukemia Virus (MM-LV) reverse transcriptase. As described in Section 5.2, negative controls for each tissue were generated, with no reverse transcriptase added to these samples. These negative controls were used in this experiment to ensure there was no contaminating genomic DNA.

For this assay, the gene-specific primers (see Section 3.2.1 and Materials and Methods) were employed to test for the presence of each of the *GhTTG* mRNAs in 0 DPA and 3 DPA Xu-142 and *fl* mutant ovules. The results were similar for all four genes tested. The sizes of the amplified bands were as expected, but no significant difference in expression between wild-type and *fl* mutants for any of the four genes tested was observed (Figure 5.3). No amplification was observed in the negative control, with no cDNA template

added, or in the non-reversed transcribed controls of DNase-treated RNA samples (data not shown).

Reverse-transcriptase PCR is only semi-quantitative and not likely to be able to detect small changes in transcript accumulation. Whilst the *in situ* hybridisation results appeared to demonstrate increased mRNA accumulation of *GhTTG1/GhTTG3* predominantly in initiating fibre cells, there was also expression in some non-fibre epidermal cells, which would be detected readily in the reverse-transcription reaction. Thus, a quantitative real-time PCR reaction was performed.

5.4.4 Real-time PCR

Real-time PCR accurately quantifies the initial amount of the mRNA in a given template, and is preferable to reverse-transcription PCR reactions that detect the amount of final amplified product at the reaction end-point. Real-time PCR analysis facilitates the exponential phase of PCR amplification to be monitored, which gives a more precise indication of mRNA levels in the template. Quantitative PCR relies on identification of the first cycle that gives a signal over the background and calculation of this threshold cycle (C_T) makes quantitation much more precise than end-point analyses (Gibson *et al.*, 1996; Heid *et al.*, 1996).

Two basic chemistries have been used in all quantitative PCR systems. One is a probe-based system, utilising a sequence-specific probe for quantitation of the template of interest. In the second method, which was used in this study, intercalating dyes fluoresce only when bound to double-stranded products generated by PCR. SYBR green dye (Applied Biosystems) was used to detect amplification in the real-time PCR analysis. SYBR green binds to the minor groove of double-stranded DNA and, as more double-stranded amplicons are produced, the SYBR green dye signal increases.

Using the 'Primer Express' program (Applied Biosystems), primers were designed to *GhTTG1* and *GhTTG3*, to detect the abundance of their mRNA in Xu-142 and Xu-142 *fl* mutant ovules. These cotton genes were selected as the most likely to be controlling fibre initiation due to the *GhTTG3* anti-sense probe demonstrating localisation to fibre initials in the *in situ* hybridisation experiment and to their ability to complement the glabrous phenotype in *Arabidopsis ttg1* mutant plants (see Chapter 4). Primers were designed to

variable regions in each of these genes, in order to distinguish between *GhTTG1* and *GhTTG3*. These primers were designed to amplify a small (approximately 100 bp) region of each gene, which is ideal for real-time analysis. Primers for real-time analysis are frequently designed to span intronic regions, so amplification from genomic DNA can be avoided. However, as none of the *GhTTG* genes contain introns (see Chapter 3), such an approach was not possible here.

The same cDNA samples used in the reverse transcriptase PCR analysis of Section 5.4.2 were used as templates in the real-time PCR assay. A control reaction was performed to ensure those samples which were not treated with a reverse transcriptase did not produce a signal during PCR amplification. Primers designed to the cotton actin gene were used as an internal control, to calibrate the reaction. The relative amounts of *GhTTG* mRNA were calculated using the comparative threshold method. 0 DPA wild-type was chosen to normalise the reaction, with the mRNA present in each of the other tissues and time-points measured as the fold difference in level, relative to actin (Appendix 2). This analysis demonstrated no significant reduction in mRNA level for either of the genes tested in the fibreless ovules compared to wild-type (Figure 5.4). In fact, *GhTTG1* expression displayed a slight increase in the fibreless ovules at both 0 and 3 DPA time-points. *GhTTG3* showed a slight decrease in expression at 0 PDA in the fibreless mutant (approximately 26% decrease in expression compared to wild-type), but mRNA levels remained unchanged at 3 DPA between the two ovule types tested (Figure 5.4). Therefore, the real-time analysis demonstrated that there was not a significant change in *GhTTG1* or *GhTTG3* transcript levels between wild-type and *fl* mutants, either at 0 or 3 DPA.

5.5 Discussion

This chapter describes an assessment of the spatial and temporal expression of the four *TTG1*-like genes in cotton. Reverse transcriptase PCR analysis demonstrated that each of the cotton *GhTTG* genes is expressed in all tissues throughout the cotton plant. This pattern of ubiquitous expression appears consistent with a general transcription profile seen for most WD-repeat genes, due to the broad diversity of roles they are known to play in eukaryotic systems. Several recent investigations of WD-repeat genes in *Arabidopsis via* RT-PCR have shown ubiquitous expression in all tissues tested (Deyholos *et al.*, 2003; Hsieh *et al.*, 2003; Shi *et al.*, 2005).

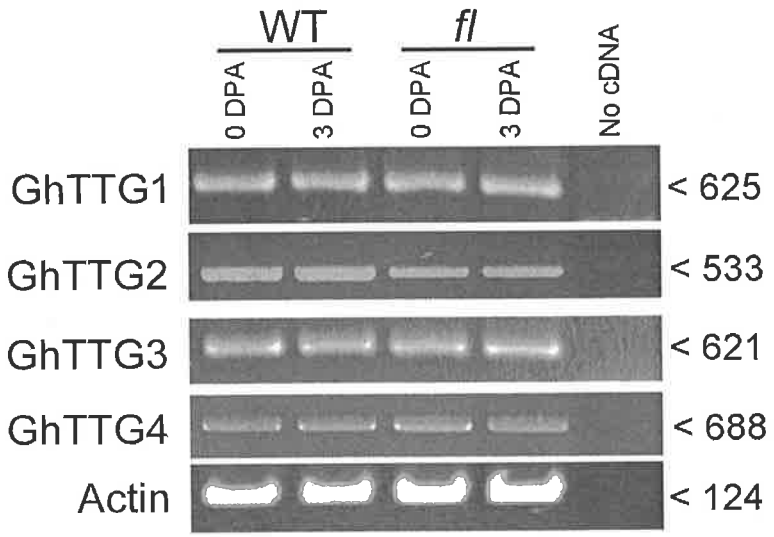


Figure 5.3: Reverse transcriptase PCR analysis of expression of cotton *TTG1*-like genes in 0 and 3 DPA ovules of Xu-142 and its fibreless mutant

Numbers on the right indicate the size of the PCR product in bp, with the target gene listed on the left. Each of the genes appears to be expressed at roughly equivalent levels in wild-type (WT) and the *fl* mutant.

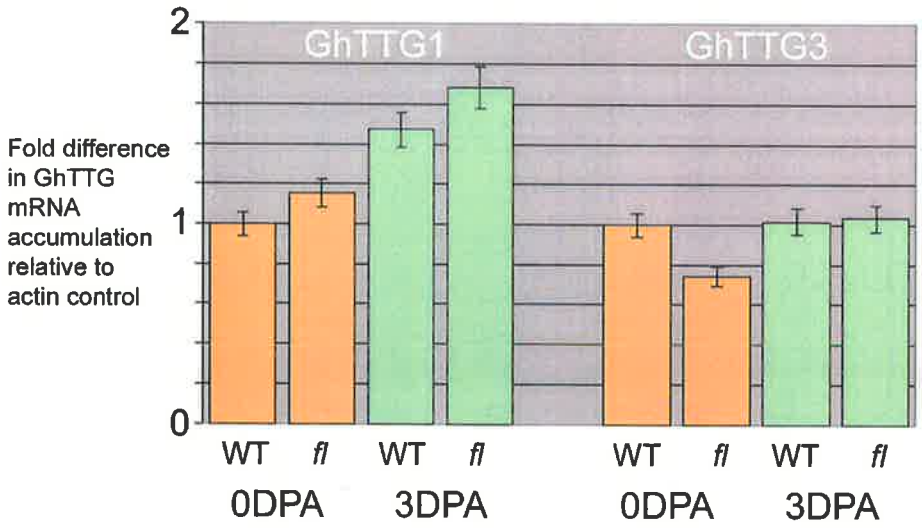


Figure 5.4: Real-time PCR analysis of *GhTTG* mRNA accumulation

Real-time PCR analysis of *GhTTG1* (left) and *GhTTG3* (right) mRNA accumulation in 0 DPA (orange) and 3 DPA (green) ovules of Xu-142 and its fibreless mutant. Relative transcript amounts were normalised with respect to cotton actin levels. For both *GhTTG1* and *GhTTG3*, expression at 0 DPA wild-type (WT) ovules was chosen as the normaliser (and set base level 1), with other levels displayed relative to this base level. Mean values were obtained from triplicate PCR amplifications. Error bars indicate the standard error of the mean.

Walker *et al.* (1999) demonstrated via Northern blot experiments that *AtTTG1* is expressed throughout all plant organs. The fact that the cotton *TTG1*-like genes could not be detected in a Northern blot is an indication that the *GhTTG* genes are expressed at a lower level than that of *AtTTG1*, and that several cotton genes, with overlapping functions, may perhaps perform the same role as the single *TTG1* gene in *Arabidopsis*. Functional redundancy is likely given the complexity of the allotetraploid cotton genome.

Reverse transcription and quantitative real-time PCR analysis demonstrated no significant difference in expression between Xu-142 and the Xu-142 fibreless mutant ovules for the *GhTTG1* and *GhTTG3* genes. This result appears in conflict with the observations of the *in situ* hybridisation analysis (Section 5.3), in which *GhTTG1/3* are expressed predominantly in those cells destined to become fibre cells. If these *in situ* hybridisation results are assumed to be accurate, the expectation would be that *GhTTG1/3* mRNA expression would be decreased in ovule samples which lack initiating fibre cells. Even the decrease observed in *GhTTG3* expression at 0 DPA (Figure 5.4) was less than that which would have been expected for a gene expressed predominantly in fibre cells, as suggested by the *in situ* hybridisation results. One possible reason for this apparent conflict is that the mutation which causes the fibreless mutation in Xu-142 acts downstream of the *GhTTG* genes, and thus these genes are still expressed at normal levels in the epidermal cells, specifically in precursor fibre cells. This is a plausible explanation, as *AtTTG1* is known to act early in the pathways it affects (reviewed in Zhang *et al.*, 2003). However, as ovules of Xu-142 are almost completely glabrous, it is likely that the mutant causing the fibreless phenotype is also acting in an early stage of fibre differentiation. Another possibility is that the mutation causing the fibreless phenotype causes a loss of spatial regulation of *GhTTG1/3*, resulting in expression throughout the ovule at a similar overall level, but not specifically in fibre cells. If this were the case, *in situ* hybridisation analysis of fibreless mutant ovules may be informative.

The similarity of *GhTTG* gene expression observed between wild-type Xu-142 and *fl* mutants is perhaps not surprising, given the investigation by Wu *et al.*, (2006), where a comparison of Xu-142 and Xu-142 *fl* 0 DPA ovules by microarray analysis identified 119 clones differentially regulated, but the *GhTTG* genes were not included in this group. One of these 119 cotton clones encodes a putative WD-40 repeat protein, but displays little sequence homology to *AtTTG1* or any of the cotton *TTG1*-like genes.

The spatial expression of other potential cotton homologues of *Arabidopsis* trichome genes has previously been investigated. Suo *et al.* (2003) performed *in situ* hybridisation on 0 and 1 DPA cotton ovules to demonstrate that the cotton MYB gene *GhMYB109* is expressed specifically in fibre initials. Unlike the WD-repeat genes studied in this chapter, *GhMYB109* is expressed at a level detectable by Northern analysis. RNA expression analysis was able to show that *GhMYB109* is highly expressed in fibre cells, and moderately expressed in ovules at 4 and 8 DPA, with little to no expression in other tissues. Additionally, it was shown that *GhMYB109* is not expressed in glabrous Xu-142 *fl* ovules at 3 DPA, in contrast to the mRNA expression observed in the wild-type ovule. In the investigation of Suo *et al.* (2003), the Northern analysis results agreed with those observed in the *in situ* hybridisation analysis, with fibre-specific expression demonstrated in both cases. This agreement is in contrast to the observations outlined in this chapter, where fibre-specific signals in the *in situ* hybridisation analysis were not confirmed by quantitative and semi-quantitative comparative PCR analysis of fibreless mutant ovules. As discussed previously, the reason for these conflicting results is unclear. In relation to the findings of Suo *et al.* (2003) it may be the case that *GhMYB109* acts later in fibre development than the Xu-142 *fl* mutation, whilst the *GhTTG1/3* genes act earlier.

As noted in Chapter 4, Wang *et al.* (2004) demonstrated that a cotton MYB gene (*GaMYB2*) had the ability to complement the *gl1* glabrous phenotype in *Arabidopsis. In situ* hybridisation analysis of this cotton gene revealed expression only in fibre initials of the cotton ovule. Real-time PCR analysis revealed low expression of the *GaMYB2* gene in the root, stem, leaves and in fibreless 6 DPA ovules. This is an example of a gene, as described for *GhTTG1* and *GhTTG3* in this chapter, which was shown by ISH to be expressed specifically in fibre cells, but is still expressed in other tissues (albeit at a low level).

The expression patterns of cotton genes detailed in this chapter and previous studies mirror the expression patterns observed in *Arabidopsis* epidermal cell differentiation. MYB genes such as *GL1* and *WER* are expressed exclusively in specific cell types, whilst *AtTTG1* is expressed throughout the plant, due to its diverse range of functions. Given their widespread expression, it could be hypothesised that, like *TTG1* in *Arabidopsis*, the

GhTTG genes are playing a number of roles in the cotton plant, possibly including fibre development, but also in leaf trichome formation or anthocyanin production.

The analysis of WD-repeat gene expression patterns has shown expression throughout all tissues, with specific transcripts observed in fibre initials in the ovule for *GhTTG1* and/or *GhTTG3*. Due to the apparent contradiction between the *in situ* hybridisation and comparative analysis of Xu-142 wild-type and fibreless ovules, it is not possible to assign a clear function for the *GhTTG* genes in cotton fibre initiation on the basis of this analysis of their expression.

Chapter 6: Molecular interactions of cotton WD-repeat proteins

6.1 Introduction

WD-repeat proteins are known to function predominantly through protein-protein interactions (Neer *et al.*, 1994; Smith *et al.*, 1999). Payne *et al.* (2000) showed through yeast two-hybrid analysis that the TTG1 protein in *Arabidopsis* interacts with the basic helix-loop-helix protein GL3, which also interacts physically with the MYB-type protein GL1. The resulting WD40-bHLH-MYB complex is conserved throughout several plant species (Ramsay and Glover, 2005). TTG1 has been shown to interact physically with other *Arabidopsis* bHLH proteins, including EGL3, TT8, and an as yet uncharacterised bHLH protein termed AtBHLH012, in combination with a variety of MYB proteins, to control various processes (see Figure 1.4, and Esch *et al.*, 2003; Baudry *et al.*, 2004; Zimmermann *et al.*, 2004). It is possible that TTG1 also interacts with the MYB protein of these WD40-bHLH-MYB complexes, but this has only been demonstrated in the case of TT2 (Baudry *et al.*, 2004). Given that protein-protein interactions are important in the function of TTG1 in trichome initiation, a yeast two-hybrid analysis was performed to investigate GhTTG interactions in cotton.

6.2 Yeast two-hybrid

The yeast two-hybrid assay is a technique for the cloning of cDNAs encoding proteins that interact with a target protein sequence (Li and Fields, 1993; Wang and Reed, 1993). This method can be used to determine whether two specific proteins interact, or to screen a library of cDNAs for an interaction with a particular protein sequence of interest.

The MATCHMAKER system (BD Biosciences) was utilised for this study, with two main expression vectors being used to create fusion proteins; pGBKT7 and pGADT7-Rec (See Appendix 1 and Materials and Methods). pGBKT7 is designed such that a sequence of interest can be expressed as a fusion protein containing the GAL4 DNA-Binding Domain (DNA- BD), whilst pGADT7-Rec expresses a protein of interest as a GAL4 Activation Domain (AD) fusion. In addition to allowing a specific protein of interest to be expressed, the pGADT7-Rec vector is supplied as a *SmaI*-linearised plasmid to promote homologous recombination of a cDNA library generated using SMART technology (see Section 6.4,

Appendix 1). Both pGBKT7 and pGADT7-Rec can replicate in *S. cerevisiae* and *E. coli* and carry markers for selection in either host. The pGBKT7 vector carries the *TRP1* nutritional marker for selection in yeast on media lacking tryptophan, whilst pGADT7-Rec contains the *LEU2* nutritional selection marker for growth on media lacking leucine. Protein interactions are detected via the reporter genes *HIS3*, *ADE2*, and *MEL1*, the expression of which is activated by the formation of a functional transcriptional activator resulting from the interaction of the two protein sequences.

Expression of the *HIS3* and *ADE2* genes can be detected by growth on media (Synthetic Dropout medium: SD) lacking histidine and adenine respectively, whilst the *MEL1* gene product can be observed, via an x-α-gal (5-Bromo-4-chloro-3-indolyl-α-D-galactopyranoside) assay, by hydrolysis of x-α-gal yielding a blue end product. The presence of three distinct reporter genes is advantageous in reducing false positives often encountered in yeast two-hybrid assays.

Section 6.3 below describes a test of specific interactions between GhTTG and known AtTTG1 interactors, whilst Section 6.4 describes a library screen of cotton fibre cDNAs for potential GhTTG interactors. The principle for these two methods is outlined in Figure 6.1.

6.3 Investigation of GhTTG interaction with GL3/EGL3

One step in determining whether the cotton WD-repeat proteins act in a similar manner to *Arabidopsis* TTG1 is to test whether they interact with the factors similar to those with which AtTTG1 is known to bind. As previously discussed, TTG1 is known to interact with the bHLH proteins GL3 and EGL3 in its role in trichome initiation. To test if any of the four cotton TTG1-like proteins are able to interact with AtGL3/EGL3, co-transformation experiments of these genes in yeast were performed.

The *GhTTG* sequences were each cloned in-frame into the pGBKT7 vector, such that they could be expressed as fusion proteins with the GAL4 DNA-BD. Similarly, the *Arabidopsis* sequences GL3 and EGL3 (provided by A. Walker) were cloned into the pGADT7 vector, allowing expression of these genes fused to the GAL4 AD.

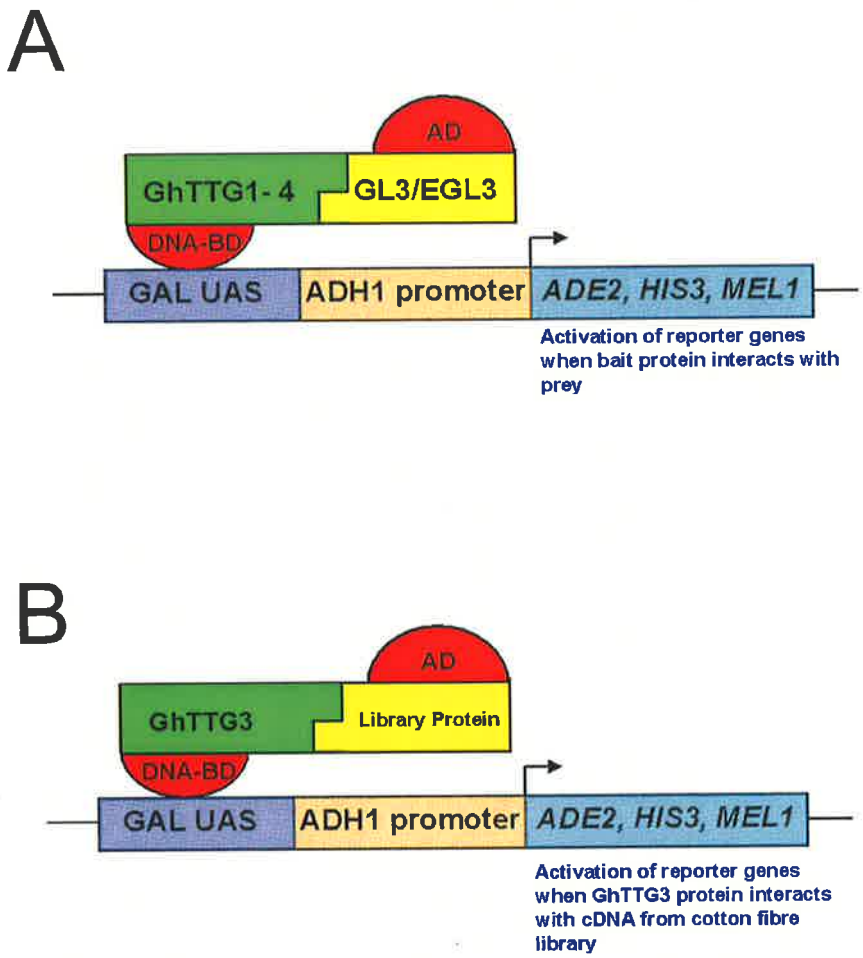


Figure 6.1: Schematic of yeast two-hybrid analysis

(A) Direct test for interaction between a cotton GhTTG protein (GhTTG1-4) and a bHLH protein known to interact with AtTTG1 (GL3 or EGL3). The GhTTG protein is fused to the GAL4 DNA-BD, and GL3/EGL3 is fused to the GAL4 AD. If an interaction occurs between the two proteins, transcription of the reporter genes is activated, whereas the reporter genes are not expressed in the absence of an interaction.

(B) Screen of a cotton fibre cDNA library for potential interactors of the GhTTG3 protein. Double-stranded cDNAs are fused with the GAL4 AD via homologous recombination in yeast, and cDNAs interacting with the bait protein (GhTTG3) result in expression of the reporter genes.

Once the constructs were generated and sequenced to verify that no errors had been introduced, and to ensure that the inserts were in the correct frame, the background level of leaky *HIS3* expression was determined. Transformants in the AH109 yeast strain show slightly elevated *HIS3* expression due to intrinsic DNA-binding properties of the bait protein such that a background level of growth is observed on SD media lacking histidine (SD/-His). 3-amino-1,2,4-triazole (3-AT) is a competitive inhibitor of the yeast *HIS3* protein, and it therefore can be used to reduce background *HIS3* in a dose-dependent manner (Struhl and Davis, 1977). To determine the levels of 3-AT required to suppress background growth, each bait plasmid was transformed into AH109 cells and plated onto a series of SD/-Trp/-His plates containing a range of 3-AT concentrations. For each of the bait plasmids, a concentration of 5 mM 3-AT was found to be sufficient to suppress background growth due to leaky *HIS3* expression (i.e. no colonies formed on these plates after five days incubation).

Each of the four cotton TTG1-like proteins was then tested for activation of the reporter genes when co-transformed into AH109 cells with GL3. As a control, AtTTG1 also was tested for an interaction with GL3; an interaction which has previously been shown to occur in yeast (Payne *et al.*, 2000). In addition, a positive control supplied with the kit was carried out involving co-transformation of pGBKT7-53 (a fusion between GAL4 DNA-BD and murine p53) and pGADT7-RecT (a fusion between GAL4 AD and SV40 Large T-antigen). p53 and SV40 Large T-antigen are known to interact strongly in the yeast two-hybrid assay (Li and Fields, 1993). The cells from each transformation were plated onto quadruple drop-out (QDO) medium (SD/-Ade/-His/-Leu/-Trp) supplemented with 5 mM 3-AT. When testing for the GhTTG-GL3 interactions, many colonies were observed on SD/-Trp/-Leu plates, indicating that both plasmids had been effectively co-transformed, but no colonies were observed on the QDO medium testing for an interaction (i.e. reporter genes *ADE2* and *HIS3* were not activated). This was the case not only for the cotton TTG sequences, but also for the AtTTG1 control. Only the supplied positive control demonstrated an interaction, with growth of colonies on QDO, which turned blue in the presence of x-α-gal (results not shown). This result demonstrated that the two-hybrid process itself was effective, but the fact that no interaction was detected between AtTTG1 and GL3 indicated there was a problem with the TTG1 and/or GL3 constructs, given that

more than one previous study has documented an interaction between AtTTG1 and AtGL3 in the yeast two-hybrid system (Payne *et al.*, 2000; Esch *et al.*, 2003; Zhang *et al.*, 2003).

Originally each of the expression constructs described was designed to include the full-length sequence of the coding regions of interest (including the start codon) downstream of the GAL4 AD or DNA-BD. To remove any complexities which may arise from the start codon of the cloned gene being utilised in preference to the GAL4 AD or DNA-BD start codon, the genes of interest were re-cloned into the pGBKT7 and pGADT7-Rec plasmids, but this time the start codon was removed from the fragment being cloned by designing PCR primers to exclude the start codon from the resulting amplicon.

This was first attempted for the AtTTG1-AtGL3 control interaction test, with the truncated (-ATG) version of AtTTG1 being cloned downstream of the GAL4 DNA-BD, and the truncated AtGL3 cloned downstream of the GAL4 AD. Co-transformations into yeast were performed with these truncated fusion proteins, but also with combinations of full-length and truncated forms.

The only assay that demonstrated a strong interaction was that in which the truncated version for both AtTTG1 and GL3 were co-transformed. Many colonies were observed for this transformation on the QDO plates and these tested positive for a *MEL1* response (blue colonies). When a full-length AtTTG1 was combined with a truncated GL3 (and vice versa), no interaction was detected, as was the case when co-expressing the two full-length sequences. This indicated that the start codons of both proteins were required to be removed before the fusion proteins could be functional in yeast, and that it was not a problem specific to a single construct.

The start codons were subsequently removed from each of the four cotton WD-repeat proteins, and new GAL4 DNA-BD fusions were generated. When each of these was tested for binding with the truncated version of AtGL3, none demonstrated an interaction, with no colonies detected on the QDO plates after 7 days incubation (Figure 6.2). In each case, control experiments confirmed that both plasmids were present in the yeast cells, and sequencing of the truncated versions of the GhTTG proteins showed the sequences were in-frame and error-free.

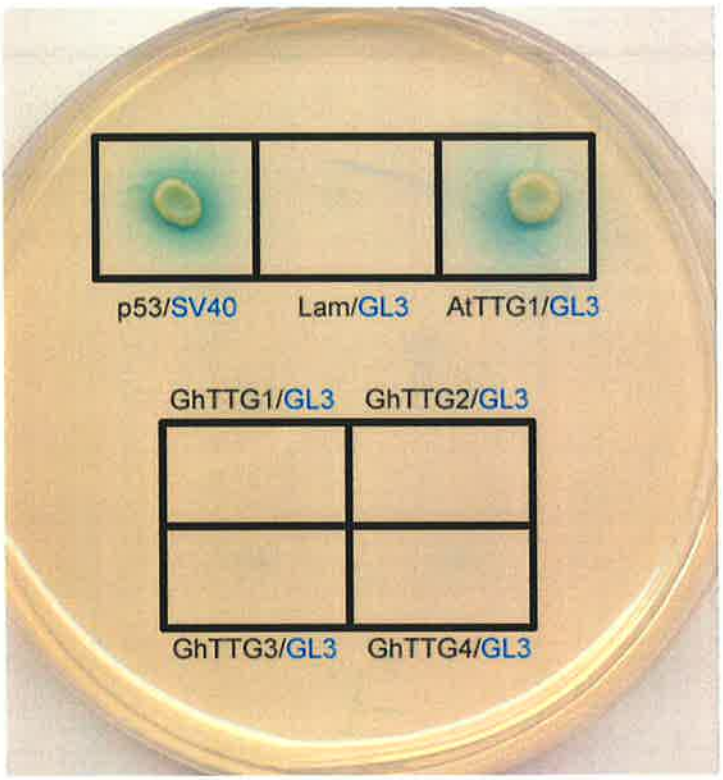


Figure 6.2: Test for interaction of GhTTG proteins with GL3

Colonies were spotted onto QDO (SD/-Ade/-His/-Leu/-Trp) with 5 mM 3-AT and x-α-gal. GL3, AtTTG1 and GhTTG proteins were expressed as fusion proteins minus the start codon. An interaction is observed for the supplied positive control (p53 and SV40) and between AtTTG1 and GL3. No interaction is detected between GL3 and the supplied negative control (pGBKT7-Lam), nor between GL3 and any of the GhTTG proteins. Black text indicates the protein fused to the GAL4 DNA-BD, whilst blue text indicates the protein fused to the GAL4 AD.

These experiments were repeated with AtEGL3 using a truncated form. Once again, a strong interaction was observed between AtTTG1 and AtEGL3 (result not shown). However, none of the GhTTG proteins showed any activation of the reporter genes when co-transformed with EGL3. The lack of interaction between GhTTG1-4 and GL3/EGL3 was somewhat surprising, given the high amino acid homology that exists between AtTTG and the GhTTG protein sequences. EGL3 = protein

Although the cotton TTG proteins did not display an interaction with AtGL3, similar bHLH proteins may be present in cotton that do interact with these sequences in cotton. There may also be other protein types that interact with the GhTTG protein to perform a function in cotton. Therefore a library screen was performed to identify potential interactors of the cotton TTG1-like proteins.

6.4 Screen for interacting proteins of GhTTG in cotton

A library screen was performed to identify putative factors interacting with cotton TTG1-like proteins. Of particular interest were members of the bHLH family of proteins that may mirror the action of GL3/EGL3 from the *Arabidopsis* system, and hence possibly be involved in cotton fibre initiation. No homologues of GL3 have been isolated from cotton, although the presumed protein sequence of a cotton clone named GhDEL65 (accession number AF336280; Matz and Burr, unpublished) displays around 40% amino acid identity to GL3.

As the tissue of primary interest in this investigation is the cotton fibre, a cDNA library was constructed using RNA extracted from 6 DPA fibres. It was expected that any cDNA involved with a cotton WD-repeat protein in the process of fibre initiation would be likely expressed in the fibre. Synthesis of the cDNA library was performed using the SMART (Switching Mechanism at the 5' end of the RNA Transcript) method (Zhu *et al.*, 2001) in which a specific MMLV reverse transcriptase is used to generate the first cDNA strand. The terminal transferase activity of this enzyme adds several C residues to the 5' end of mature mRNAs during first strand synthesis. Second strand synthesis is then primed by an oligonucleotide that binds to the C residues to promote template switching and synthesis of the second cDNA strand by MMLV. Since the MMLV enzyme does not typically add C residues to the 5' end of truncated mRNAs, full-length cDNAs are preferentially synthesised. Hence, SMART cDNA libraries normally contain a higher ~~frequency~~ proportion of full-

length cDNAs than conventionally-generated cDNA libraries (Sugahara *et al.*, 2001; Zhu *et al.*, 2001).

The library screen is performed by constructing a plasmid library consisting of cDNA from a tissue of interest (in this case cotton fibre) fused to a transcriptional activation domain from a yeast protein (GAL4). Yeast cells containing one or more reporter genes regulated by a yeast minimal promoter (which allows basal transcription only) along with the target protein sequence are co-transformed with the library cDNA clones. If a cDNA encodes a protein that interacts with the target protein sequence, the fusion protein will bind to the target and the AD will activate reporter gene expression. cDNAs encoding protein domains that interact with the target protein sequence may be isolated from yeast clones that are positive for reporter gene expression.

The GhTTG3 protein was chosen as the "bait" for this library screen, due to its high similarity to *Arabidopsis* TTG1, and the fact it has been shown to complement the mutant phenotype in *Arabidopsis ttg1* mutants (Chapter 4). The library screen was attempted using the truncated version of the GhTTG3 bait protein (i.e. without the start codon, Section 6.3). The fact that the full-length AtTTG1 was unable to activate the reporter genes when combined with AtGL3 in this system (while the truncated version was able to do so) suggested that the truncated version of GhTTG3 may be more suitable for identifying potential interactors of GhTTG3 in this library screen. The GhTTG3:DNA-BD construct was co-transformed into yeast with double-stranded cotton fibre cDNA and the linearised pGADT7-Rec vector, which allows for expression of the cDNAs as AD fusions via homologous recombination in yeast (see Appendix 1). The transformed cells were spread on QDO media containing SD -Leu/-Trp/-Ade/-His medium supplemented with 5 mM 3-AT. A 1:100 dilution of the transformed cells also was plated on SD –Leu/-Trp medium to enable estimation of the library size and transformation efficiency. The plates were then incubated at 30°C for five days. Approximately 570,000 clones were screened in this assay, as indicated by the number of colonies growing from the 1:100 dilution of the transformed cells plated on SD/-Trp/-Leu medium.

This library screen produced 15 colonies on the QDO medium (5 mM 3-AT). These clones were re-streaked twice on the same medium, but with x-α-galactosidase added to test for

expression of the *MEL1* reporter gene, to confirm their *ADE2*$^{+}$*HIS3*$^{+}$*MEL1*$^{+}$ phenotype. Upon re-streaking, and testing for activation of the *MEL1* gene, three of the fifteen clones appeared to activate each of the reporter genes strongly. Those that did not were assumed to be false positives (result not shown). The plasmid from each of the three positive clones was isolated, transferred into *E. coli*, purified and then digested with appropriate restriction enzymes to confirm the presence of the pGADT7-Rec vector carrying a cDNA insert. The cDNA inserts were approximately 400, 900 and 1900 bp in length (Figure 6.3). The isolated plasmids were propagated in yeast with the truncated GhTTG3:DNA-BD plasmid, to confirm the activation of the reporter genes (Figure 6.4). Transformation of yeast with the three isolated plasmids alone (not shown) or with the empty pGBKT7 vector (Figure 6.4A) demonstrated no growth, indicating that the cDNAs did not activate the yeast reporter genes independently of the GhTTG3 sequence. A quantitative binding assay measuring x-α-gal activity (see Materials and Methods) confirmed the interaction of each of three clones with GhTTG3. This assay also indicated that Clone #2 was the strongest of the three interactors, whilst Clone #3 was the weakest (Figure 6.4B; Appendix 3).

The three plasmids were sequenced to determine the identity of the cDNA insert. The sequences were then analysed using BLASTN and TBLASTX (Altschul *et al.*, 1990; Altschul *et al.*, 1997) to identify any similar protein-coding sequences in the sequence databases. BLASTN searches of each of the cDNA inserts revealed similar but not identical ESTs from *G. hirsutum* and *G. raimondii* for the three clones (Table 6.1).

TBLASTX conceptually translates the query nucleotide sequence in all six reading frames and compares the amino acid sequence with a similar translation of database sequences. The TBLASTX analysis demonstrated that the Clone #1 cDNA encoded a putative NAC-domain protein, Clone #2 a TCP-like protein, whilst Clone #3 demonstrated no significant homology to a sequence in the database (Table 6.1).

Clone #:	1	2	3
Size of cDNA insert	846 bp	1752 bp	249 bp
Size of truncated ORF	317 amino acids	210 amino acids	29 amino acids
BLASTN search of cotton ESTs in Genbank	Gossypium hirsutum fibre cDNA clone 497359, 1e-131	Gossypium raimondii cDNA clone GR__Eb10I06 5', mRNA sequence, 7e-89 Gossypium arboreum 7-10 dpa fibre library cDNA clone GA__Ed0005G12f, mRNA sequence, 2e-83	Gossypium raimondii cDNA clone GR__Ea43P19, 1e-5
Putative identity of protein encoded by cDNA insert	NAC-domain protein	TCP-domain protein	Ubiquitin hydrolase
Highest scoring TBLASTX results	•*Arabidopsis thaliana* transcription factor ANAC103; NM_125802; 9e-10 • *Arabidopsis thaliana* transcription factor NAC2; NM_120523; 5e-7	• *Oryza sativa* Transcription factor PCF3-like; Q6K7P2; 4e-40 • *Arabidopsis thaliana* Auxin-induced bHLH transcription factor; Q9C518; 1e-21	• *Arabidopsis thaliana* T01490 ubiquitin hydrolase homolog; T01490; 2e-4

Table 6.1 Sequence analysis of cDNA insert from three clones isolated in a yeast two-hybrid library screen. Highest scoring results from BLASTN and TBLASTX searches of full-length insert are shown. TBLASTX results include species name, gene name, Genbank accession number and e-value score.

Each of the cDNA inserts appeared to be partial sequences truncated at their 5' ends, as they lacked a 5' untranslated region (UTR) or a start codon with a consensus translation initiation sequence, and each was shorter than candidate homologues from other plant species.

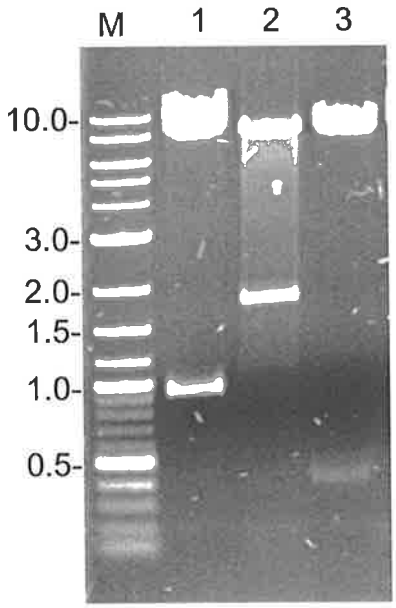
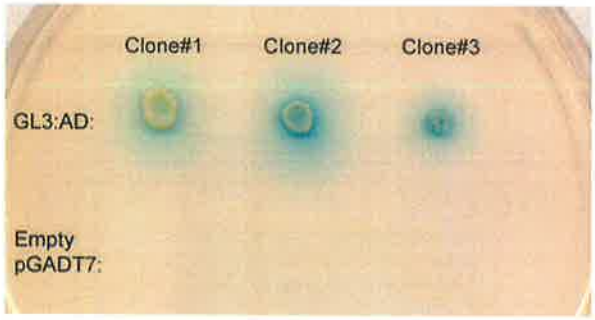


Figure 6.3: Restriction analysis of cDNAs isolated in yeast two-hybrid library screen

*Eco*RI/*Bam*HI digestion of the three plasmids isolated in library screen (Clones 1-3). M= NEB 2-log DNA marker (marker bands are every 100 bp from 0.1 to 1.0 kb- sizes are indicated in kb). The vector backbone is represented by the 10 kb fragment, whilst the lower band represents the cDNA insert. Insert sizes range from 400 bp to 1.9 kb, with each "insert band" containing approximately 50 bp of vector sequence.

A



B

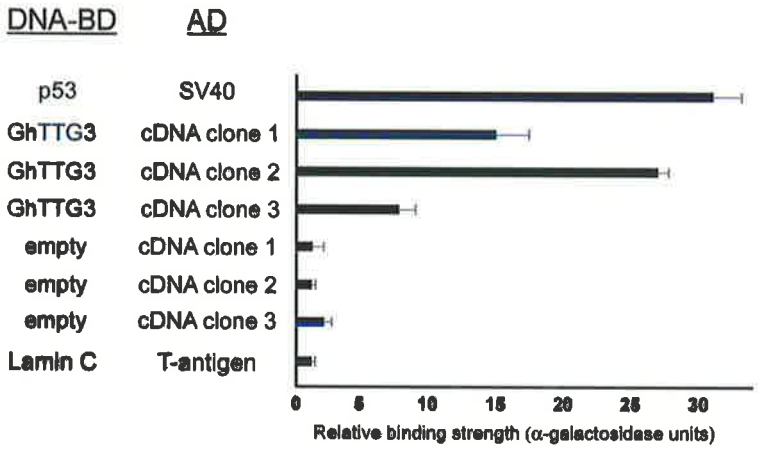


Figure 6.4: Test of interactors of GhTTG3 isolated in library screen

(A) Plasmids isolated from library screen co-transformed with truncated GhTTG3 (top row) or empty pGBKT7 (second row), spotted onto QDO (SD/-Ade/-His/-Leu/-Trp) with 5 mM 3-AT and x-α-gal. An interaction is demonstrated by a blue patch of growth.

(B) X-α-Gal Quantitative Binding Assay. Relative binding strength of the three cDNA clones to either GhTTG3 or the empty pGBKT7, as measured by quantitative x-α-gal assay. The supplied positive (p53 and SV40) and negative (Lamin C and T-antigen) controls are included for comparison. Values are a mean of triplicate assays, and standard deviations are shown.

Clone #3 contains a small cDNA insert of only 249 bp (Figure 6.5C), and was the lowest in strength of the three interactors (Figure 6.4B). It had no significant homology to any other plant protein sequence in the database showing only low (29%) identity to a short fragment of a ubiquitin hydrolase. As ubiquitin and related proteins are commonly identified as false-positives in yeast-two hybrid library screens (Serebriiskii *et al.*, 2000), Clone #3 was not studied in great detail. A discussion of the two remaining clones follows. The sequence derived from Clone #1 is referred to as GhNAC, whilst the Clone #2 cDNA insert is termed GhTCP.

6.5 Interacting Clone #1: NAC domain protein (GhNAC)

Clone #1 displayed the highest homology to members of the NAC-domain family (Table 6.1). The NAC gene family is named after its three founding members; NAM (*no apical meristem* of petunia), ATAF1/2 and CUC2 (*CUP-SHAPED COTYLEDON 2* of *Arabidopsis*). Genes containing the NAC domain are plant-specific, and an analysis of *Arabidopsis* transcription factors reveals more than one hundred members of this gene family (Riechmann *et al.*, 2000). NAC-domain proteins have been implicated in a wide variety of processes, including plant defence, cell elongation, stress response and lateral organ development (reviewed by Olsen *et al.*, 2005). Exactly how NAC-domain proteins exert their specific function is largely unknown.

Numerous NAC-domain proteins have been demonstrated to be nuclear-localised, supporting the supposition that they are transcription factors (Xie *et al.*, 2000; Greve *et al.*, 2003; Fujita *et al.*, 2004; Robertson, 2004; Taoka *et al.*, 2004). Despite this, no typical NLS has been identified, although Kikuchi *et al.* (2000) proposed a putative NLS within the NAC-domain of family members in rice. The NAC-domain typically spans approximately 150 amino acids in the N-terminal region of the protein, generally beginning around 25 amino acids from the start. The C-terminal region is very diverse among NAC family proteins. Clone #1 contains an ORF of 217 amino acids, the N-terminal 27 amino acids of which are part of the consensus NAC-domain (Figure 6.5A). This region of GhNAC is very similar to that of NAC2 of *Arabidopsis* (He *et al.*, 2005b), whilst the C-terminal region does not show significant homology to any sequence in the database. A 3'UTR of approximately 200 bp is also present in Clone #1 (Figure 6.5A).

Ooka *et al.* (2003) analysed the predicted sequences of 75 NAC-domain proteins from rice and 105 from *Arabidopsis*. A phylogenetic approach was used to classify each of the proteins based on the sequence of their NAC-domain, resulting in 18 sub-groups being identified. The GhNAC cDNA sequence appears to be most closely related to the "NAC2" sub-group of this study, which contains the *Arabidopsis* NAC2 protein and several other uncharacterised *Arabidopsis* and rice presumed NAC-domain proteins. The highest scoring sequences in a BLASTX search with GhNAC all derive from this sub-group. None of the members has an assigned function, although it has been suggested NAC2 may play some role in lateral root formation (He *et al.*, 2005). It is thus difficult to suggest how GhNAC might function in cotton. No NAC-domain genes previously have been characterised in cotton, although two *G. raimondii* ESTs (CO123587 and TC30282; Udall *et al.*, 2006) show significant homology (approximately 50% identity) to GhNAC (Table 6.1).

Many NAC-domain proteins contain a transactivation (TA) domain at the C-terminus, which has been shown to activate expression of genes in yeast (Duval *et al.*, 2002; Taoka *et al.*, 2004; He *et al.*, 2005b; Oh *et al.*, 2005). Previous studies investigating NAC-domain interactions via yeast two-hybrid screens have used a C-terminal truncated version of the protein to eliminate the transactivation domain (Xie *et al.*, 2000). Therefore it is possible that the retrieval of a NAC-domain protein in this assay is due to activation of the yeast reporter genes by this TA-domain, rather than by a genuine interaction with GhTTG3 (i.e. it is a false positive). However, GhNAC does not show significant homology in the C-terminal region to any of the proteins reported previously to cause activation of yeast reporter genes. In addition, the presence of Clone #1 on its own and with the empty pGBKT7 vector did not activate the reporter gene expression in the two-hybrid assay (Figure 6.4). However, further investigation to verify the interaction is required.

6.6 Interacting Clone #2: TCP-domain protein (GhTCP)

Clone #2 demonstrates the highest sequence similarity to PCF3 from rice, in addition to several other genes from the TCP family. This is a group of transcription factors that forms a small sub-group of the wider basic helix-loop-helix family.

The TCP family, which is plant-specific, was named after its first-characterised members; teosinte branched 1 (*tb1*) from maize, cycloidea (*cyc*) from snapdragon, and PCF1 from rice (Cubas *et al.*, 1999). The *cyc* gene in *Antirrhinum* is involved in the control of floral

symmetry (Carpenter and Coen, 1990), whilst *tb1* affects the fate of maize axillary meristems (Doebley *et al.*, 1995). Cubas *et al.* (1999) investigated the biological roles of the proteins encoded by *cyc* and *tb1* by studying regions conserved between the two amino acid sequences that may act as functional domains, as the processes controlled by the two genes shared many common themes. One of the conserved regions was found to form a basic helix-loop-helix structure, although it was unrelated in primary amino acid sequence to the other bHLH transcription factors. This bHLH domain was identified as being closely related to that found in two rice proteins, PCF1 and PCF2 (Kosugi and Ohashi, 1997) and this family of proteins was termed TCP. The family currently contains over 100 known genes, of which 25 are from *Arabidopsis* (Reeves and Olmstead, 2003). 58-60 amino acids comprise the bHLH domain which is thought to be involved in DNA-binding and protein-protein interactions. The domain is strongly conserved between members of the TCP family, although a high level of diversity is observed outside of this bHLH domain. The proteins of this family fall into two sub-classes; one of which is more closely related to CYC and TB1 (termed TCP-C), whilst the other more closely resembles the PCFs (termed TCP-P) (Cubas *et al.*, 1999). The main feature that distinguishes the two classes is a bipartite NLS found in TCP-C proteins, of which only a portion is present in the TCP-P group.

The GhTCP sequence retrieved in this yeast two-hybrid assay aligns more closely with the TPC-P subgroup, with the most similar sequences all coming from this sub-group. Alignments show that the basic helix-loop-helix region of the TCP domain present in GhTCP is highly conserved between species, particularly within the TCP-P subgroup (Figure 6.6). GhTCP contains six key residues (i.e., --R--KV--R-RR--) known to be conserved in many basic helix-loop-helix transcription factors (Littlewood and Evan, 1998).

The GhTCP cDNA insert contains the C-terminal 210 amino acids of the presumed protein (including the entire TCP bHLH domain), with approximately 1 kb of 3' UTR (Figure 6.5B). Most TCP-like bHLH proteins are between 300 and 600 amino acids in length, so it is unknown exactly how much of the full-length GhTCP protein is missing from the cDNA insert, although the most closely related sequences are approximately 310 amino acids in length. The TCP domain, as defined by Cubas *et al.* (1999) is present from bases 34 to 198 of the 1.7 kb cDNA insert (Figure 6.5B). The most closely related cotton sequence to

GhTCP is a cDNA termed GhBHLH, identified as an auxin-induced basic helix-loop-helix transcription factor (Jaradat and Allen, 1999). The role of GhBHLH in cotton is unknown, but it is ubiquitously expressed (Jaradat and Allen, 1999). GhTCP is also similar to the cotton ESTs AI730418 and AI726525 (Blewitt *et al.*, unpublished), an indication that it may be a member of a multigene family, as suggested by Jaradat *et al.* (1999).

6.7 Discussion

6.7.1 WD40 interactions

WD40 proteins are known to interact with a number of different protein types in the various roles that they play in plant development. Of particular interest in this study is the interaction of AtTTG1 with the bHLH proteins GL3 and EGL3 during trichome initiation, and it was hypothesised that similar proteins may exist in cotton that bind the cotton TTG1 homologues.

Several examples of WD-repeat proteins interacting with bHLH proteins have been documented. A yeast two hybrid assay using the Perilla WD-repeat protein PFWD demonstrated an interaction with a Perilla bHLH transcription factor, which in turn was shown to interact with a MYB protein (Sompornpailin *et al.*, 2002). PFWD was also shown to function cooperatively with the Perilla bHLH protein in onion cells, leading to the recruitment of PFWD from the cytoplasm into the nucleus. In *Z. mays* a genetic interaction between the WD40 protein PAC1 (which affects anthocyanin production) and a maize bHLH anthocyanin regulator (IN1) has been demonstrated by double mutant analysis (Carey *et al.*, 2004). Analysis of anthocyanin production in petunia has uncovered two key proteins, AN1 (bHLH) and AN11 (WD40), which demonstrate high sequence similarity to GL3 and TTG1 respectively. It is believed that these two proteins, along with the MYB-type protein AN2, form a complex involved in anthocyanin biosynthesis (de Vetten *et al.*, 1997). In an investigation of anthocyanin production in the species Japanese morning glory, a MYB (InMYB1) and WD-repeat protein (InWDR1) were isolated, and are hypothesised to act coordinately with a bHLH transcription factor (InbHLH2) to activate structural genes for anthocyanin biosynthesis, although a physical interaction has not yet been investigated (Morita *et al.*, 2006).

Additional WD-repeat proteins, with little sequence similarity to TTG1, have also been shown to interact with bHLH proteins. These include the *Arabidopsis* COP1 protein, which

Figure 6.5: Sequence analysis of cotton cDNAs isolated in two-hybrid screen

Conceptual translation is shown beneath cDNA sequence, with stop codon underlined in red. In each case the ORF shown is in the same frame as the GAL4 AD in the pGADT7-Rec vector.

(A): Sequence of GhNAC cDNA insert of Clone #1 Blue underline= NAC-domain

```
   TATAGGCTTGAAGAAAAAATATTGGCTGACAGAGGGCCTGTTCAAGATATGTGTATGCTT
 1 ---------+---------+---------+---------+---------+---------+ 60
   Y  R  L  E  E  K  I  L  A  D  R  G  P  V  Q  D  M  C  M  L

   TATGTTGTTTTCAAGAAGGATGGTGTGGGCCCAAGAAATGGTTCCCAATACGGGGCCCCA
61 ---------+---------+---------+---------+---------+---------+ 120
   Y  V  V  F  K  K  D  G  V  G  P  R  N  G  S  Q  Y  G  A  P

   TTCAAGGAAGAGGAGTGGAGTGACGACGAGCAGGCAGATGTGCCTGGTGCTGGTTCCCTT
121 ---------+---------+---------+---------+---------+---------+ 180
   F  K  E  E  E  W  S  D  D  E  Q  A  D  V  P  G  A  G  S  L

   TGCGGCCTGTCTACAGTAGCTGTTGGTGGTGATAATAGCTCCTGTGTCCCCGAGAGTGTA
181 ---------+---------+---------+---------+---------+---------+ 240
   C  G  L  S  T  V  A  V  G  G  D  N  S  S  C  V  P  E  S  V

   TGTGCTGGGCCTTCTGTCGAGTCAAGTGGGTGGTGTACACAACCATCGATGGGTCATACT
241 ---------+---------+---------+---------+---------+---------+ 300
   C  A  G  P  S  V  E  S  S  G  W  C  T  Q  P  S  M  G  H  T

   AATGGCGATGTTAATGTCACCACTGCCATTGATACTAATGCTTATGCTGCCCCTACATGC
301 ---------+---------+---------+---------+---------+---------+ 360
   N  G  D  V  N  V  T  T  A  I  D  T  N  A  Y  A  A  P  T  C

   GCGGATGCCCCTGAAGTTCCAGTGGTCGCTGTTGCTGAGGATGTTCCTACATTGTTGGAG
361 ---------+---------+---------+---------+---------+---------+ 420
   A  D  A  P  E  V  P  V  V  A  V  A  E  D  V  P  T  L  L  E

   GCGCCTCAACCTGCTCTGGTTCCTCAAGATCAAGAATCGAATGATAACATGTTATCTATG
421 ---------+---------+---------+---------+---------+---------+ 480
   A  P  Q  P  A  L  V  P  Q  D  Q  E  S  N  D  N  M  L  S  M

   TTGGAGTGCTTCTTTGAAGAGGGTGCATTTGGTACACATGAAGGTCAGAGGAGGGGCAAT
481 ---------+---------+---------+---------+---------+---------+ 540
   L  E  C  F  F  E  E  G  A  F  G  T  H  E  G  Q  R  R  G  N

   GCAACTGATTCTTCCTTCAAGGATACTCCTGTTTCTGATGAGGATGATATTTCAATCTTG
541 ---------+---------+---------+---------+---------+---------+ 600
   A  T  D  S  S  F  K  D  T  P  V  S  D  E  D  D  I  S  I  L

   TTGGCCTCTTTTACTGATGTCGACAATTCAAATAACTGGTCTTATCCTTTTTAATAAAGT
601 ---------+---------+---------+---------+---------+---------+ 660
   L  A  S  F  T  D  V  D  N  S  N  N  W  S  Y  P  F  *

   AATTGTTTCAACTATTGGAACTGGAAGCAATGTTTTTAGATTACTTGTTTTATTCTTAGG
661 ---------+---------+---------+---------+---------+---------+ 720

   AATAAATGCTGTTGTTTGGTTTACCAGTGTCATTGACTTGTTTCTATTTGATTATCATGT
721 ---------+---------+---------+---------+---------+---------+ 780

   TGTGCTCTCTATACATTTTCTTTAATGGAATGTCTGGCATATTATTTGTGAAAAAAAAAA
781 ---------+---------+---------+---------+---------+---------+ 840

   AAAAAA
841 ------ 846
```

(B): GhTCP cDNA insert of Clone #2: Blue underline = TCP bHLH domain

```
    GGGAACGGTACCCTTGCCGTCAAGAAACCTCCCTCCAAAGACCGTCACAGCAAAGTAGAC
 1  ---------+---------+---------+---------+---------+---------+ 60
    G  N  G  T  L  A  V  K  K  P  P  S  K  D  R  H  S  K  V  D

    GGTCGTGGACGGAGGATAAGGATGCCCATCATTTGCGCCGCTCGTGTTTTCCAGTTGACT
61  ---------+---------+---------+---------+---------+---------+ 120
    G  R  G  R  R  I  R  M  P  I  I  C  A  A  R  V  F  Q  L  T

    CGCGAGTTGGGTCATAAATCCGACGGTCAGACCATCGAGTGGCTGTTACGCCAGGCCGAG
121 ---------+---------+---------+---------+---------+---------+ 180
    R  E  L  G  H  K  S  D  G  Q  T  I  E  W  L  L  R  Q  A  E

    CCTTCCATCATAGCCGCTACAGGTACGGGGACGACTCCAGCCAGCTTCTCCACCGTGTCA
181 ---------+---------+---------+---------+---------+---------+ 240
    P  S  I  I  A  A  T  G  T  G  T  T  P  A  S  F  S  T  V  S

    GTTTCCGTCCGCGGAGCCAATTCCACTTCTTTGCCGTCAACGACTTCTTCGTCGTTGGAT
241 ---------+---------+---------+---------+---------+---------+ 300
    V  S  V  R  G  A  N  S  T  S  L  P  S  T  T  S  S  S  L  D

    CATAAGCCTTTGTTGGGTTCCACCCCCTTTTTACTCGGAAAACGTGTCCGGCCCGACGAC
301 ---------+---------+---------+---------+---------+---------+ 360
    H  K  P  L  L  G  S  T  P  F  L  L  G  K  R  V  R  P  D  D

    GATAACGCCGGGAAGGACGATTCCGGAGGGGCCACGGTGGGACCGGGCGTGGGTTCTATC
361 ---------+---------+---------+---------+---------+---------+ 420
    D  N  A  G  K  D  D  S  G  G  A  T  V  G  P  G  V  G  S  I

    GTGGGGCCCGCTTGAACACGGGTGCGTTCTGGGCTGTTCCAGCTAGGCCTGACTTTGGTC
421 ---------+---------+---------+---------+---------+---------+ 480
    V  G  P  A  G  T  R  V  R  S  G  L  F  Q  L  G  L  T  L  V

    AAATATGGAGCTTCGCGCCCCCGCCTCCTCCCGAGATGGTGGTACAAACAGCGGCTCAGC
481 ---------+---------+---------+---------+---------+---------+ 540
    K  Y  G  A  S  R  P  R  L  L  P  R  W  W  Y  K  Q  R  L  S

    AACCAGCTGCAGCTGCGTTTTTCGTTCAGCAACAGCAAGCGATGGGTGAAGCGTCGGCGG
541 ---------+---------+---------+---------+---------+---------+ 600
    N  Q  L  Q  L  R  F  S  F  S  N  S  K  R  W  V  K  R  R  R

    CGAGGGTTGGCAATTATCTTCCCGGTCATCTGAATTTGTTAGCCTCCTTGTCGGGTGCTC
601 ---------+---------+---------+---------+---------+---------+ 660
    R  G  L  A  I  I  F  P  V  I  *

    CGGGAGGTTCGGGTCAAAGAGACGATGACCCTCGTTGAATTTGGGTGGACCGGTTTACCC
661 ---------+---------+---------+---------+---------+---------+ 720

    TTGTTGTTGGGAGTATATATATATTTATATGAATACATAGATTCTATAAATATTTGTGTA
721 ---------+---------+---------+---------+---------+---------+ 780

    TCAATGTTGTTTTCATGTGTGTGGATGTGGAAGATCATTTCTAGTTGGAATTTTTAGGGA
781 ---------+---------+---------+---------+---------+---------+ 840

    TAGCCCACTTAGGGTTTTTGAGGAAACGATTTGGGAGTTTGGAGTCTGATGAATGGTTTC
841 ---------+---------+---------+---------+---------+---------+ 900

    AAATTAGGGTTTTCCCCTTAGCACTTCTATTTGATGATTAATTAGGGTTTTAGAGCTTTC
901 ---------+---------+---------+---------+---------+---------+ 960


    ACAATTGTTTAGGAATTTGAGAGTTTATGTTGGAGAAGAAGAGGGGCTCAAGAATTAGAG
961 ---------+---------+---------+---------+---------+---------+ 1020
```

```
     CTTTCATGAGTGATATATACTTGGGGATTTCTTTTGTCATTAATTTGGGTTTGCTTATTG
1021 ---------+---------+---------+---------+---------+---------+ 1080

     CTCTTTGATTTTGAGCAAGTACTTGCTGATTTCTTGCTGCATTGAAGAATTAGAGGTTTC
1081 ---------+---------+---------+---------+---------+---------+ 1140

     AAGCCTGGATTTGGAGAATTCAGGCTTTTCCCTCATTGGCAAGTTCTTCCATTGTTCTTG
1141 ---------+---------+---------+---------+---------+---------+ 1200

     TCATAGGAACGACGAACGAGCAAAGCGATTAATCTTACTTCTGGATTGTACTTGATTTAC
1201 ---------+---------+---------+---------+---------+---------+ 1260

     ATATTATTGGGGTTTGATCGGTTAACCAAAATAGGGTGACTGGATGCTGAACgCCAGAGA
1261 ---------+---------+---------+---------+---------+---------+ 1320

     TTTGCTTGCTTGGCTACTTAACTTGATGCTGtGGATATATGGTCTCCCTAAATATTCTGG
1321 ---------+---------+---------+---------+---------+---------+ 1380

     AATTCCTTCTATTGCTATGGTTGTTTGAGGACTATTAGTTTCTTTGCAGCTGTTGCTTAG
1381 ---------+---------+---------+---------+---------+---------+ 1440

     CATCGGAAATTATGGTTCTTAAGTGCATAGCTCATGGTATGCTGTTGACGTGGGTTGATC
1441 ---------+---------+---------+---------+---------+---------+ 1500

     TGGCCCTCTCTCTTCTGCTTTCAGTGTTTGTGTTTGATCGCATGGTCTCAAAGGAGACTG
1501 ---------+---------+---------+---------+---------+---------+ 1560

     CTGTGTGAAAGATTgCATTCCTTTCTGTAAGATTTTTATTTTGGAAAACTTTAGCTGTAC
1561 ---------+---------+---------+---------+---------+---------+ 1620

     TTTTAGCCATTTCTTAGATTTATCGTGTGCAAATAATGTGTTCTTCATGTATTGTTTTGG
1621 ---------+---------+---------+---------+---------+---------+ 1680

     GTGCCTGTTGTGTGAATAAAAAAACCCATATGTTTTTCTCAAAAAAAAAAAAAAAAAAA
1681 ---------+---------+---------+---------+---------+---------+ 1740


     AAAAAAAAAAA
1741 ----------+--1752
```

(C): cDNA insert of Clone #3

```
    GGTCCAAAAAGGGAACTTCGTTTGATGTCTCGAGATCGGATCAGGATTGTATGTGCTTAT
  1 ---------+---------+---------+---------+---------+---------+ 60
    G  P  K  R  E  L  R  L  M  S  R  D  R  I  R  I  V  C  A  Y

    GTTCTTCCTTATACTATAGTAGCTAACTGAGCTTGTTTGTGTATTAGGGTTACCAAGTTA
 61 ---------+---------+---------+---------+---------+---------+ 120
    V  L  P  Y  T  I  V  A  N  *

    ATCTCTCTAGCATTTGAGATTTGCTATAAACTGTGGAGGACATGTTCTTGTATCTGTTGT
121 ---------+---------+---------+---------+---------+---------+ 180

    TATGTTATTGAAATGCATCATTTTGTTGCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA
181 ---------+---------+---------+---------+---------+---------+ 240

    AAAAAAAAA
241 --------- 249
```

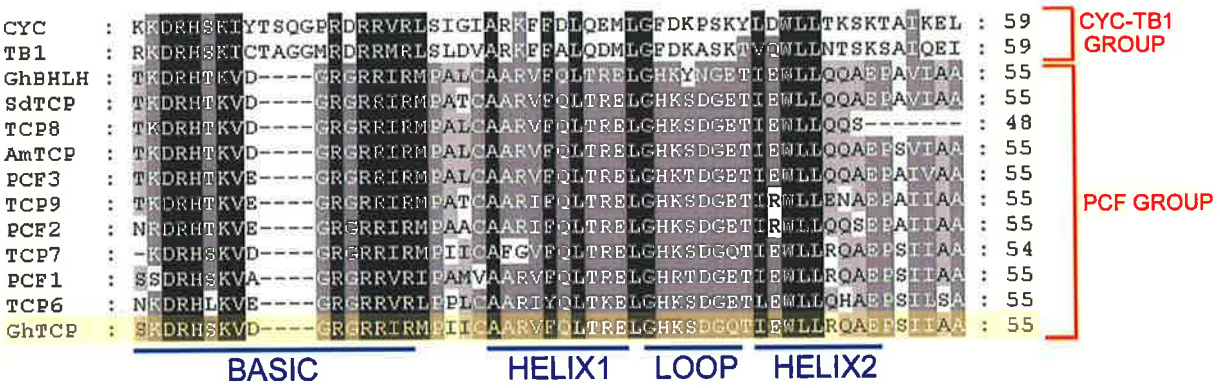


Figure 6.6: Alignment of selected TCP-domain proteins

An alignment of the bHLH domains of a selection of sequences from the PCF-like sub-group, including the cotton sequences GhTCP (this study, highlighted in yellow) and GhBHLH (Jaradat and Allen, 1999), plus the founding members of the TP1/CYC subgroup for comparison.

has been shown to interact in a yeast two-hybrid assay with HFR1, a bHLH transcription factor (Jang *et al.*, 2005). The *Drosophila* WD-repeat protein Groucho was found to interact with Hairy, a bHLH protein involved in repression of transcription, in addition to other hairy-related bHLH proteins, Deadpan and Enhancer of split (Paroush *et al.*, 1994). Similarly, the human orthologue of Groucho, TLE, is known to interact with a bHLH factor Hes1, also forming a transcriptional repression complex (Grbavec and Stifani, 1996).

An interaction between AtTTG1 and GL3 was verified in this study, but the role that the TTG1 protein plays as part of the WD-40/bHLH/MYB complex is not entirely clear. One hypothesis is that TTG1 plays a role in mediating the transfer of interacting factors from the cytoplasm to the nucleus, although evidence suggests that it may in fact be a transcription factor such as GL3 which recruits TTG1 to the nucleus (A. Walker, pers comm.). TTG1 may simply function to stabilise the MYB-bHLH transcription complex, raising the question as to whether TTG1 plays an active role, or acts purely as a bridge between transcription factors.

6.7.2 Testing of GhTTG protein interactions

It was expected that the cotton TTG1-like proteins, in particular GhTTG1 and GhTTG3, would show an interaction with AtGL3 in yeast because they were able to complement the *ttg1* mutant phenotype (Chapter 4). It is presumed that GhTTG1 and GhTTG3 substitute for AtTTG1 in the transformed plants, and hence interact with *Arabidopsis* protein factors, presumably GL3, to perform TTG1 functions. It is perhaps the case that the over-expression in the complementation experiments (the GhTTG genes were driven by the 35S promoter) caused an interaction to occur. However, at the level of expression monitored in the yeast assay, no interaction could be detected. It is also possible that the GhTTG proteins introduced into the *ttg1* mutants were able to interact with a different but similar bHLH protein from *Arabidopsis* (e.g. AtBHLH012) and form a complex to promote trichome initiation. Another explanation could be simply that in yeast the GhTTG proteins do not act in their normal manner, although this appears unlikely given the highly similar AtTTG1 was able to demonstrate a strong interaction with GL3 in this study. Western blot analysis to confirm that the GhTTG proteins were indeed being expressed may have been informative. Also, fusing the GhTTG genes with the AD domain, and testing for an interaction with GL3 fused to the DNA-BD domain may have yielded results, although

most published reports of TTG1 interactions have TTG1 fused to the DNA-BD (Payne *et al.*, 2000; Esch *et al.*, 2003)

The fact that GhTTG proteins do not interact with GL3 (whilst the AtTTG1 protein does) suggests that the GhTTG protein sequences may provide clues to the key residues involved in the protein interaction between GL3 and AtTTG1, given the high similarity between GhTTG and AtTTG1. As expected, the WD-repeat regions have been identified as being required for an interaction but a more detailed search for critical residues may now be possible.

It is peculiar that the start codon was required to be removed from the AtTTG1 and AtGL3 coding regions before they could demonstrate an interaction in the yeast two-hybrid experiment. Previous studies demonstrating an interaction between these two proteins have used the full-length sequences with no difficulties (Payne *et al.*, 2000; Esch *et al.*, 2003). Whilst the expression vectors used were different, the same GAL4 system was utilised, with the BD and AD domains fused upstream of the proteins of interest. It is assumed that the TTG1 or GL3 start codon was being utilised in preference to the DNA-BD or AD start codon when the full-length insert was present, although why this might be occurring is unknown.

6.7.3 A cotton fibre cDNA library screen for interactors of GhTTG3

The library screen for interactors of GhTTG3 in cotton fibre cDNA yielded three primary clones. The retrieval of a cDNA (GhTCP) encoding a putative protein sequence that formed a basic helix-loop-helix structure was of most interest, given previous published reports of TTG1 interaction with several bHLH proteins, in particular with GL3 during its role in trichome initiation.

It is tempting to speculate that GhTCP is a binding partner of GhTTG3 *in vivo*, and that it is through this interaction that GhTTG3 exerts some developmental effects, perhaps including fibre growth. For a similar complex to form as the trichome-inducing complex of TTG1-GL3-GL1 in *Arabidopsis*, a MYB partner would need to be identified. Previous studies have shown that some members of the TCP family interact with MYB-type proteins. In snapdragon two pairs of transcription factors, two MYB proteins (RAD and DIV) and two TCP proteins (CYC and DICH), are known to control floral asymmetry. It is

thought CYC and DICH influence petal development by activation of RAD via physical interaction (Corley *et al.*, 2005). Also in snapdragon, *CINCINNATA* (*CIN*) is a TCP-type protein which interacts with MIXTA in its role in cell differentiation (Crawford *et al.*, 2004). *MIXTA* encodes a R2R3 MYB-type protein that promotes conical cell differentiation in the petal epidermis. MIXTA has also been shown to promote trichome initiation when over-expressed in tobacco (Glover *et al.*, 1998). A highly similar MYB protein (termed AmMYB MIXTA-like1; AmMYBML1) has since been found to control trichome formation in snapdragon, specifically in the ventral petal tube (Perez-Rodriguez *et al.*, 2005). It is not known whether CIN also interacts with this protein.

The interaction of TCP-domain proteins with MYB-type proteins may be an indication of some functional similarity to other bHLH proteins that also interact with MYB proteins. In snapdragon, MIXTA is related to GL1 from *Arabidopsis*, and the fact that CIN and GL3 proteins interact with MIXTA and GL1 respectively may indicate some form of conserved function between TCP and canonical bHLH proteins. This study suggests that the cotton TCP interacts with a WD-repeat protein (GhTTG3). It is therefore possible that a form of the WD40-bHLH-MYB complex which controls anthocyanin and trichome initiation in other plant species also is present in cotton. Whether such a complex could be functioning in fibre initiation, in anthocyanin production or in leaf trichome initiation, is unknown. It is necessary to isolate the full-length cotton TCP sequence, to analyse its expression pattern to gain some information of its function and to look for additional interactors of GhTCP in cotton. Given the interaction of the MIXTA protein with a TCP domain protein in snapdragon, a potential interactor of GhTCP in cotton is the protein encoded by GhMyb25, a gene isolated by Wu *et al.* (2006) in a search for genes involved in cotton fibre initiation. GhMyb25 demonstrates expression profiles consistent with a role in cotton fibre initiation, and encodes an R2R3 MYB protein which is closely related to the MIXTA and MIXTA-like (AmMYBML1) proteins present in snapdragon. In addition, Yang *et al.*, (2006) isolated two fibre-specific homologues of AmMIXTA/AmMYBML1 in cotton with high transcript abundance at 0 and 3 DPA, which are down-regulated in a fibreless mutant. MIXTA-like transcription factors could therefore represent possible MYB protein partners of GhTCP, as part of a complex involving a cotton WD-repeat protein.

Some caution must be exercised in making links between this TCP-domain protein and regular bHLH proteins. The significance of discovering bHLH protein is unknown,

especially given there is no direct evidence for a conservation of function between the TCP group of bHLH proteins and the canonical bHLH proteins of which GL3, EGL3 and TT8 are members. It is not known whether the similarity to regular basic helix-loop-helix domains in TCP proteins is restricted purely to secondary structure, or whether it also extends to function (Heim *et al.*, 2003). In addition, it has not been precisely determined which region of the GL3, ELG3 and TT8 bHLH proteins interact with TTG1, and hence it may even be a region outside of the bHLH domain responsible for the interaction, as each of these proteins show a level of conservation outside the bHLH domain. It is thought that the MYB GL1 protein binds to the N-terminal end of the GL3 protein, whilst TTG1 appears to interact at the C-terminal end, which contains the bHLH motif (Payne *et al.*, 2000), but the precise binding region has not been determined.

6.8 Summary

The analysis of the interactions of cotton TTG1-like proteins has shown that the GhTTG proteins differ in protein binding activity from AtTTG1, in that they do not demonstrate an interaction with bHLH proteins GL3 and EGL3 in the yeast two-hybrid system. This may reveal a difference in the way the TTG1-like proteins from the different species perform their function, or perhaps reflect some variation in their behaviour in the yeast system. cDNAs encoding putative NAC-domain and TCP-domain proteins have been identified as interacting with the GhTTG3 protein, although further research is required to verify these interactions, and evaluate their possible roles in cotton. The isolation of a bHLH domain protein raises the possibility that similar protein complexes involving TTG1 may form in both *Arabidopsis* and cotton.

Chapter 7: Discussion and conclusions

7.1 Summary of results

The molecular mechanisms that underpin the initiation of cotton fibres have remained particularly elusive. A better understanding of the genetic control of fibre initiation and development will allow for more targeted strategies for fibre improvement through genetic engineering technologies, as well as providing novel plant germplasm for conventional plant breeding.

The aim of this study was to analyse cotton genes displaying high sequence homology to the *TTG1* gene, which plays a vital role in trichome initiation in *Arabidopsis*. The premise behind this work is the idea that *Arabidopsis* trichomes and the fibres of cotton are similar structures and may therefore have some molecular properties in common. Four such genes had been isolated previously in this laboratory (*GhTTG1-4*, S. Orford), each of which encodes a putative protein product with four highly conserved WD-repeat regions. GhTTG1 and GhTTG3 demonstrated the highest amino acid identity (79% and 80% respectively) to TTG1, whilst GhTTG2 and GhTTG4 displayed a lower identity (63% and 62%).

The genomic organisation of the *GhTTG* genes was examined. Given the high sequence similarity existing between *GhTTG1* and *GhTTG3*, and between *GhTTG2* and *GhTTG4*, it was originally hypothesised that these genes may represent allo-alleles, with each gene of the closely related pairs originating from a different ancestral diploid genome. However, this was not the case, as genomic PCR analysis revealed that all four *TTG1*-like genes isolated in this study are evolutionarily derived from the D genome of *G. raimondii*, and therefore do not form two homoeologous pairs. Genomic Southern analysis using probes specific to each cotton WD-repeat gene demonstrated that the *GhTTG* genes are single or low-copy number, although there appears to be additional related family members, based on the observed cross-hybridisation of non-specific probes.

The four genes were tested for functional homology to *TTG1* via cross-species complementation experiments. Previous to this investigation, two of the four genes

(*GhTTG1* and *GhTTG3*) were shown to complement the anthocyanin-deficient phenotype of the *Matthiola incana ttg1* mutant (Humphries *et al.*, 2005). The study described here utilised a stable transformation to demonstrate that these same two genes are able to promote trichome initiation in glabrous *Arabidopsis ttg1* mutants. A closer examination of *GhTTG3* transgenic plants showed that *GhTTG3* was able to complement the full range of *ttg1* mutant phenotypes. This gives an indication that the protein products of these cotton WD-repeat genes act in a similar manner to the AtTTG1 protein, and may hence play a role in trichome or cotton fibre initiation.

To further assess the roles that the *GhTTG* genes may be playing, their pattern of expression was analysed in cotton. RT-PCR analysis showed that each of the four cotton WD-repeat genes was expressed ubiquitously throughout the cotton plant, and at all stages of fibre development. Although no significant change in mRNA expression of the *GhTTG* genes was detected between ovules of wild-type and a lintless variety in a RT-PCR investigation, an analysis of spatial mRNA expression via *in situ* hybridisation within the wild-type cotton ovule suggested that *GhTTG1*, *GhTTG3* or perhaps both genes are expressed predominantly in initiating fibre cells, a pattern consistent with a role in promoting fibre growth.

As WD-repeat proteins are to exert their function primarily through interactions with other proteins, the molecular interactions involving GhTTG proteins in cotton were investigated by way of a yeast two-hybrid assay. None of the four TTG1-like cotton proteins were able to interact with the *Arabidopsis* bHLH proteins GL3 and EGL3, despite the fact that the TTG1 protein can readily do so. A screen of a 6 DPA fibre cDNA library uncovered a NAC-domain and TCP-domain protein as potential interactors of GhTTG3, the latter of which, interestingly, forms a bHLH secondary structure.

7.2 Further experiments

Whilst many of the aims of this project have been successfully accomplished, more experiments are required to complete some of these findings, and to ultimately address the question of the role of *GhTTG* genes in cotton.

An extensive real time PCR analysis of a range of lintless mutant cotton ovules could lead to a more confident assertion as to whether the *GhTTG* genes are indeed expressed

predominantly in fibre cells. *In situ* hybridisation analysis of the ovules of lintless varieties would also be of interest, to compare with the results observed in the wild-type ovule.

Further work is required to verify the interactions observed in the yeast two-hybrid library screen, such as coimmunoprecipitation or GST-pull down experiments. In addition, 5' RACE analysis would also be required to isolate the full-length cDNAs of the potential interactors, as only partial sequences were obtained in the cDNA library screen. To test whether any of the interacting transcription factors have an effect on the nuclear localisation of the GhTTG protein, a co-bombardment experiment into onion cells could be also be performed. Particle bombardment experiments performed in this laboratory (Chalmers, 2004) demonstrated that the GhTTG proteins are localised in the cytoplasm of onion cells. It may therefore be of interest to determine if any of the transcription factors identified in the yeast-two hybrid screen are able to translocate the GhTTG protein to the nucleus via co-bombardment experiments. A similar experiment by Sompornpailin *et al.* (2002) demonstrated that the *Perilla frutenscens* WD-repeat protein PFWD is localised to the cytoplasm in onion cells, but can be imported into the nucleus upon co-expression with a bHLH protein. It would also be of interest to perform a second yeast two-hybrid library screen, using the GhTCP bHLH protein as bait, to determine whether there exists any MYB-type proteins in cotton (such as GhMYB25) that interact with this protein, and perhaps forming a complex with similarity to the WD40-bHLH-MYB trichome promoting complex observed in *Arabidopsis* and other plant species.

Although complicated by the tetraploid nature of the genome, the generation of transgenic cotton is a vital step to establish whether the cotton WD-repeat genes analysed in this study play any role in cotton fibre development. The *GhTTG* gene of interest could be over-expressed in cotton, or inhibited by way of antisense or RNAi suppression. An analysis of these transgenic cottons, and comparison of fibre initiation and traits with wild-type, would shed light on the role that the *GhTTG* genes are playing. Whilst a suppression experiment could produce a more pronounced effect on fibre development if the targeted gene is indeed involved in fibre initiation (i.e. leading to no or reduced fibre production), such an approach is susceptible to the effects of redundant gene function in the tetraploid genome. At the time of their publication Wang *et al.* (2004) were in the process of performing such a gene suppression experiment to test the role of the potential *GL1* homologue *GaMYB2* in cotton, although no results have been published as yet. An over-expression experiment

may be less affected by the presence of similar genes in the cotton genome, although any effect that the WD-repeat genes might be exerting on fibre initiation would need to be dose-dependent to be observed. Indeed, over-expression of *TTG1* has been shown to have no effect on trichome numbers or morphology in *Arabidopsis*, even though *TTG1* is required for trichome initiation (Payne *et al.*, 2000). Conversely, over-expression of *Perilla* PFWD in wild-type *Arabidopsis* has been shown to result in an increase in trichome numbers (Sompornpailin *et al.*, 2002), and thus it is possible that an effect could be observed upon the up-regulation of a *TTG1*-like gene in cotton. Whether this effect would be on fibre development, leaf trichomes, anthocyanin production or some other pathway remains an important question.

7.3 Arabidopsis trichomes as a model system for cotton fibre development

The model plant *Arabidopsis* has been widely examined, with a view to extrapolating data for use in commercially significant crop species. *Arabidopsis* trichomes are an excellent model system to study all aspects of cell differentiation including cell fate determination, cell cycle regulation, cell polarity and cell expansion. A number of researchers have used this system as a tool to study cotton fibre development, specifically in aspects of fibre initiation. However, the link between *Arabidopsis* trichomes and cotton fibres remains to be fully established. Wu *et al.*, (2006) found that the initiation of leaf and stem trichomes is not affected in a number of fibreless cotton mutants, giving an indication that there exist distinct processes controlling the regulation of the two structures. As yet no homologues of trichome regulators have been shown to play a role in cotton fibre initiation, although several candidate *GL1* homologues have been studied in cotton, with some showing promising expression patterns (Suo *et al.*, 2003; Wang *et al.*, 2004).

Whilst the ability of two of the cotton *TTG1*-like genes to functionally replace *TTG1* in its role in trichome initiation is an indicator that they may be playing some role in cotton fibre initiation, it is certainly a result to be interpreted with caution. Several cases exist where a WD-repeat gene has been shown to perform a function when transformed into a heterologous species, a role in which the gene does not participate in the source species (Lloyd *et al.*, 1992; de Vetten *et al.*, 1997; Carey *et al.*, 2004). The results of this study leave open the possibility that the genetic regulators of trichome and cotton fibre initiation may be conserved between species, but more information regarding the role of *GhTTG* genes in fibre development must be obtained.

Results from the yeast two-hybrid library screen (Chapter 6) indicate that a basic helix-loop-helix factor of the TCP-domain family interacts with the WD-repeat protein GhTTG3 in cotton. It could be argued that this situation mirrors that of the trichome-initiating complex in *Arabidopsis*, in which TTG1 interacts with the bHLH protein GL3. The third member of this complex in *Arabidopsis* is GL1, a MYB transcription factor which interacts directly with GL3. As several TCP-domain proteins have been shown to interact with MYB-type proteins (most significantly the trichome-related MIXTA protein from snapdragon), it is possible that a version of the WD40-bHLH-MYB complex that controls trichome initiation in *Arabidopsis* also exists in cotton, containing GhTCP and a MIXTA-related MYB protein.

Evidence arguing against the function of such a complex in cotton fibre initiation is the notion that the transcriptional machinery that control trichome (or fibre) development in the Rosids (including *Arabidopsis* and cotton) differs from that which regulates trichome formation in the Asterids (including snapdragon, tobacco, and petunia) (Serna and Martin, 2006). This notion is largely based on a phylogenetic analysis of *MYB*-type genes involved in trichome initiation in a range of species, revealing that the *MYB* genes controlling trichome (or fibre) development in cotton and *Arabidopsis* are likely to be closely related to *GL1*, whilst those involved in trichome initiation in the Asterids are more closely related to the *MIXTA* gene. Additionally, over-expression of *MIXTA* induces the production of supernumerary trichomes in other Asterid species (Payne et al., 1999; Glover *et al.*, 2004), but has no effect in *Arabidopsis* (Payne *et al.*, 1999), strengthening the idea that trichomes of the Asterids and Rosids have developed independently, by way of convergent evolution. Thus, it appears more likely that a *GL1*-related *MYB* gene, as opposed to a *MIXTA*-related *MYB* gene, would be involved in fibre initiation in cotton. This may not necessarily be the case, given that no *GL1*-like genes have been shown definitively to play a role in cotton fibre initiation, and that the most promising candidate gene isolated by Wu *et al.* (2006) in a microarray screen for cotton fibre initiation genes was the *MIXTA*-related *GhMyb25*. In addition, Yang *et al.*, (2006) isolated two putative MIXTA homologues in cotton which a potential role in fibre initiation. MIXTA-related MYB proteins do not contain the amino acid signature required for interaction with R-like bHLH proteins (such as GL3 or EGL3) (Grotewold *et al.*, 2000), however interactions with TCP-domain bHLH proteins have been established (Crawford *et al.*, 2004; Corley *et al.*, 2005), and as such it is

conceivable that a complex involving a TTG1-like WD-repeat protein, a TCP-domain bHLH factor, and a MIXTA-like MYB protein may be involved in the process of cotton fibre initiation. The elucidation of the genes involved in trichome initiation in a wider array of species will be important in increasing our knowledge of the trichome evolution, and hence determining their suitability for use as model systems in cotton fibre research.

7.4 Implications for cotton fibre development

The primary aim of this study was to obtain a better understanding as to whether the *GhTTG* genes are playing a role in cotton fibre initiation. An issue arising from this research is how this information can be used to improve cotton fibre yields, if indeed the *GhTTG* genes are functioning in the development of the fibre. Given that *AtTTG1* plays its role in the initiation component of trichome development, any *GhTTG*-driven increase in yield would likely be brought about by increasing the proportion of epidermal cells of the cotton ovule which initiate into fibre cells. An increase in yield may be achievable by generating transgenic cotton with one of the *GhTTG* genes up-regulated (either via a constitutive or fibre-specific promoter). If fibre initiation were increased in such a transgenic plant, the genotypes may be used in breeding programs in combination with other desirable traits, such as herbicide or insect resistance, to produce an economically improved field crop. However, a yield increase may instead require the up-regulation of a transcription factor that acts cooperatively with a GhTTG protein (as is the case for *Arabidopsis* trichomes, where over-expression of *GL3*, but not *TTG1*, has been shown to increase trichome numbers). Such a transcription factor may be the GhTCP protein isolated in the yeast two-hybrid screen, or another factor involved in a fibre cell initiation complex. An unknown factor is the metabolic penalty that could arise as a result of increasing the number of fibres per ovule.

It is possible that the roles of the *GhTTG* genes are restricted to those displayed by *TTG1* in *Arabidopsis*, such as leaf trichome initiation and anthocyanin production. Even if the *GhTTG* genes are not playing a role in fibre initiation, it is perhaps conceivable that the over-expression of one of these genes, or even of *AtTTG1*, in cotton may promote a higher level of fibre initiation. Previous studies have demonstrated that the transformation of *Arabidopsis* genes into cotton can confer specific traits of commercial benefit, such as stress tolerance (Yan *et al.*, 2004) and increased photosynthetic performance, leading to increased fibre production in salty conditions (He *et al.*, 2005a). However, if it is not the

role of the native *GhTTG* genes to control fibre initiation, it is unlikely that the required molecular machinery will be present to allow over-expression of the gene to have an effect on fibre cell differentiation in the ovule epidermis.

Whilst trichomes are known to provide a defence against insect attack, it is unlikely an increase in leaf trichome numbers in cotton will be advantageous in regards to fibre yield. In a recent study, the *GL3* gene was over-expressed in the crop species canola (*Brassica Napus),* with a view towards creating an insect-resistant crop with a dense covering of trichomes (Gruber *et al.*, 2006). Indeed, *GL3*-overexpressing canola plants exhibited up to 1000-fold increase in leaf trichome numbers, and preliminary experiments suggest this layering perturbs insect attack. However, such a trichome-based defence is most useful at the early stages of growth, for example at seedling emergence, where the canola plant is most vulnerable to feeding of the flea-beetle. In the case of cotton, particularly in Australia, the most damage is inflicted by pests targeting the cotton boll in advanced stages of cotton plant growth (Matthews and Tunstall, 1994), at which point leaf trichomes provide little, if any, defence. In fact, an analysis of a cotton genotype, pilose, with a high leaf trichome density demonstrated a higher incidence of the cotton flea-hopper and their eggs on leaves, in comparison to genotypes with normal trichome density (Mekala *et al.*, 2003). Thus, it appears unlikely any manipulation of the *GhTTG* genes would aid in improving fibre yield if *GhTTG* function were restricted to leaf trichome formation in cotton.

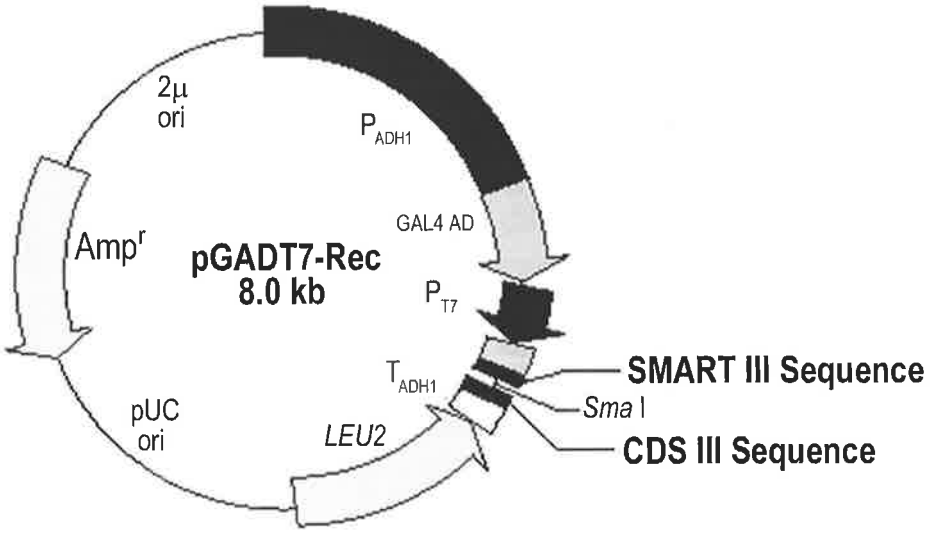
7.5 Conclusion

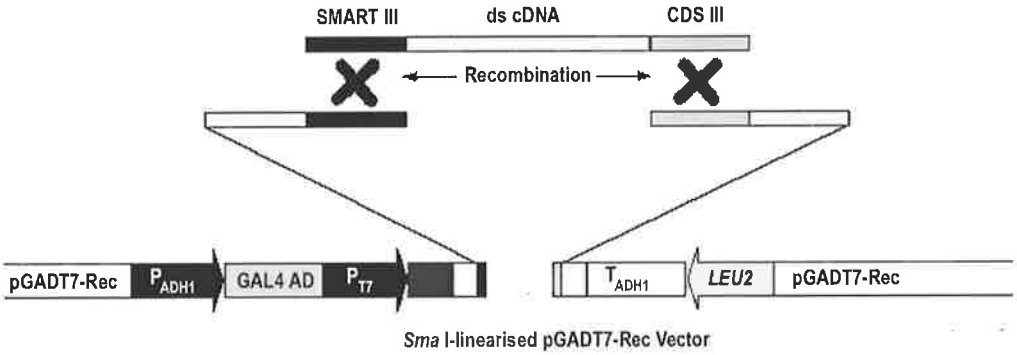
This study analysed four cotton genes with high similarity to *TTG1* in *Arabidopsis*, in order to gain more information regarding their potential involvement in the initiation of cotton fibres. This research has also provided some insight into the relationship between the epidermal cell differentiation in cotton and *Arabidopsis*. The results from cross-species complementation experiments strengthened the notion that two of the *GhTTG* genes may play a pivotal role in cotton fibre development, and the expression of these genes predominantly within fibre cells of the cotton ovule provides further evidence for such a function. It remains to be determined whether a similar transcriptional regulatory network which controls trichome initiation in *Arabidopsis* is present in cotton, and whether such a complex might be regulating fibre initiation. The continued use of model systems to direct future research, combined with the rapid increase in the availability of genomic data from

cotton, should enable the elucidation of the genes involved in fibre initiation, with subsequent potential benefits to the Australian cotton industry.

Appendix 1: Vector Maps

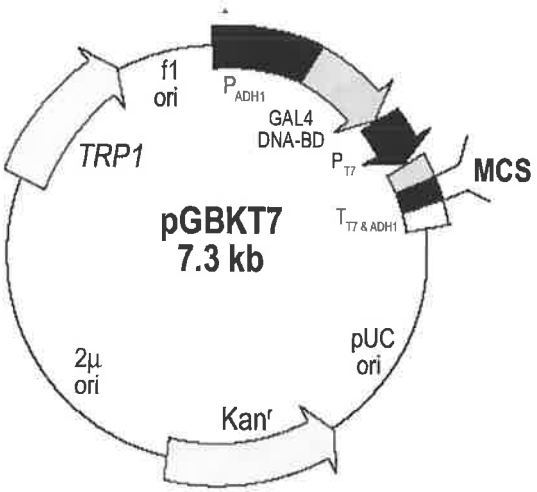
pGADT7-Rec





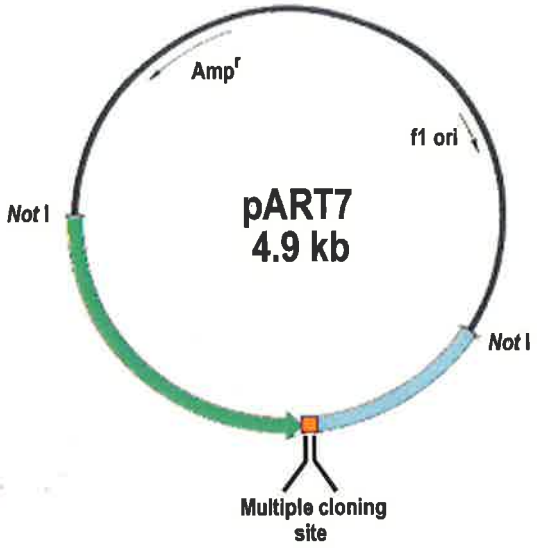
The pGADT7-Rec plasmid (top) and schematic of the process of cloning cDNA into pGADT7-Rec by homologous recombination using SMART technology *in vitro* (bottom). Reproduced from BD Biosciences pGADT7-Rec vector information booklet PT3530-5.

pGBKT7

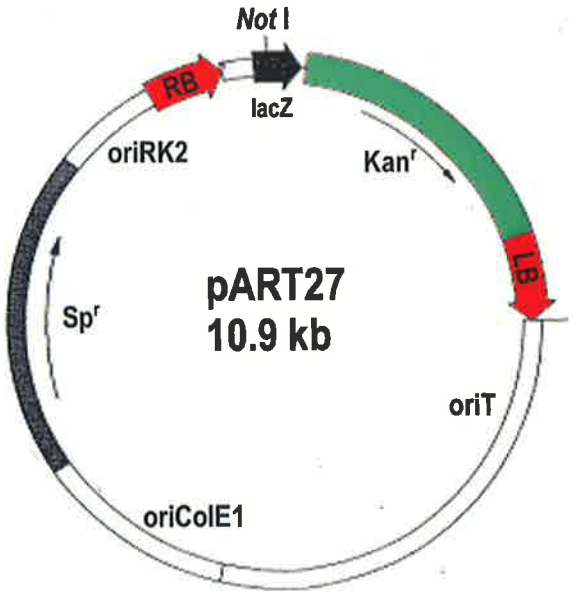


The pGBKT7 plasmid, reproduced from BD Biosciences pGBKT7 vector information booklet PT3248-5.

pART7



pART27



The pART7 and pART27 vectors, reproduced from Gleave, 1992. The gene of interest is cloned into the multiple cloning site of pART7, and the *NotI* fragment of pART7 can then be transferred to the *NotI* site of pART27 that allows for replication in *E. coli* and *Agrobacterium*.

Appendix 2: Real-time PCR data

Expression levels were determined using the Comparative threshold method. Data presented are the mean value of triplicate reactions, with standard deviations shown. Expression at 0 DPA of the wild-type ovule was used to normalise *GhTTG* transcript levels in comparison to the internal control actin. Fold difference refers to change in *GhTTG* transcript levels compared to actin levels, relative to the normaliser (0 DPA WT sample). Calculations were performed as described in the booklet "Guide to Performing Quantitation of Gene Expression using Real-time Quantitative PCR" (Applied Biosystems).

GhTTG1:

	GhTTG1 C_T	Actin C_T	ΔCT	ΔΔCT	Fold difference
0 WT	23.64 ± 0.03	18.90 ± 0.28	4.74 ± 0.28	0.00 ± 0.28	1
0 *fl*	23.53 ± 0.03	19.00 ± 0.19	4.53 ± 0.19	-0.21 ± 0.19	1.15
3 WT	22.20 ± 0.18	18.02 ± 0.36	4.18 ± 0.40	-0.56 ± 0.40	1.47
3 *fl*	22.20 ± 0.08	18.21± 0.05	3.99 ± 0.09	-0.75 ± 0.09	1.68

GhTTG3:

	GhTTG3 C_T	Actin C_T	ΔCT	ΔΔCT	Fold difference
0 WT	24.44 ± 0.06	18.90 ± 0.28	5.54 ± 0.29	0.00 ± 0.29	1
0 *fl*	25.08 ± 0.10	19.00 ± 0.19	6.08 ± 0.19	0.54 ± 0.19	0.74
3 WT	23.54 ± 0.08	18.02 ± 0.36	5.52 ± 0.37	-0.02 ± 0.40	1.01
3 *fl*	23.71 ± 0.09	18.21± 0.05	5.50 ± 0.10	-0.04 ± 0.09	1.03

Appendix 3: α-galactosidase assay

The α–galactosidase quantitative assay was performed as described in the BD Biosciences Yeast Protocols Handbook, with results shown in the table below. Each of the three clone isolated in the yeast two-hybrid library screen were co-transformed with GhTTG3 and the empty vector. Interaction strength was compared co-transformation with the empty pGBKT7 vector, and the kit-supplied positive (P53 and SV40) and negative (T-antigen/Lam) controls. Briefly, triplicate overnight cultures grown from colonies containing two plasmids (Column1 in table below) were measured for optical density at 600nm. After performing the required steps for the 1 ml α–galactosidase assay (using 1.5ml cuvettes) OD_{410} was measured. These values were used in the equation below, and the mean of the triplicate reactions was recorded (final column in table below).

Replicate #	1		2		3		α-galactosidase units (average)
OD (nm)	600	410	600	410	600	410	
P53 / SV40 (+)	0.852	0.217	0.884	0.232	0.834	0.208	31.24 ± 3.12
#1 / GhTTG3	0.763	0.092	0.802	0.103	0.795	0.085	14.50 ± 3.22
#2 / GhTTG3	0.944	0.204	0.855	0.198	0.823	0.182	27.26 ± 1.56
#3 / GhTTG3	0.734	0.055	0.676	0.044	0.722	0.051	8.58 ± 1.34
#1/ empty pGBKT7	0.656	0.010	0.723	0.008	0.588	0.005	1.41 ± 1.02
#2/ empty pGBKT7	0.702	0.008	0.666	0.007	0.676	0.007	1.31 ± 0.58
#3/ empty pGBKT7	0.688	0.012	0.851	0.014	0.821	0.015	2.12 ± 0.85
T / Lam (-)	0.657	0.007	0.634	0.009	0.781	0.005	1.27 ± 0.34

Raw data for quantification of α-galactosidase units for each interaction tested, using the equation:

$$\alpha\text{-galactosidase units} = OD_{410} \times V_f \times 1000 / (\varepsilon \times b) \times t \times V_i \times OD_{600}$$

Where V_f = Final volume of assay = 992 µl

V_i = Volume of culture medium supernatant added = 8 µl

t = elapsed time of incubation = 60 minutes

ε = ρ-nitrophenol molar absorbtivity at 410nm = 16.9 (ml/µmol)

b = light path = 1 cm

Appendix 4: Sequence identification numbers

Data for the *TTG1*-like cotton sequences were deposited in the DDBJ/EMBL/GenBank database with accession numbers AF530907 (GhTTG1 cDNA), AF530908 (*GhTTG1*), AF530909 (GhTTG2 cDNA), AF530910 (*GhTTG2*), AF530911 (*GhTTG3*) and AF530912 (*GhTTG4*).

TIGR Gene Indices and Genbank Accession numbers of WD-repeat sequences listed in Figure 3.5

Rice1: TC279520
Barley1: TC133641
Maize1: AY115485
Aquilegia1: TC14746
Grape1: TC47485
Perilla1: AB059642
Petunia1: U94748
Tomato1: TC164085
Potato: TC125946
JMG1: AB232779
CMG1: AB232771
Poplar1: TC22493
Apple: AF220203
Iceplant: TC5440
*Arabidopsis*1: TC185139
Matthiola1: AJ586861
Medicago1: TC105711
Soybean1: TC216015
JMG2: AB232780
CMG2: AB232778
Arabipdosis2: U94746
*Arabidopsis*3: X94788
Aquilegia2: TC8356
Grape2: TC40216
Soybean2: TC216013
Medicago2: TC97113
Wheat: TC270954
Barley2: TC148204
Sorghum: TC93164
Maize2: AY339884
Rice2: TC66884

John A. Humphries, Amanda R. Walker, Jeremy N. Timmis and Sharon J. Orford (2005) Two WD-repeat genes from cotton are functional homologues of the Arabidopsis thaliana TRANSPARENT TESTA GLABRA1 (TTG1) gene.
Plant Molecular Biology, v. 57 (1), pp. 67-81, January 2005

References

Adams, K. L., R. Cronn, R. Percifield and J. F. Wendel (2003). "Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing." Proceedings of the National Academy of Sciences of the United States of America **100**(8): 4649-4654.

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**(17): 3389-402.

Arabidopsis Genome Initiative (2000). "Analysis of the genome sequence of the flowering plant *Arabidopsis* thaliana." Nature **408**(6814): 796-815.

Arpat, A. B., M. Waugh, J. P. Sullivan, M. Gonzales, D. Frisch, D. Main, T. Wood, A. Leslie, R. A. Wing and T. A. Wilkins (2004). "Functional genomics of cell elongation in developing cotton fibers." Plant Molecular Biology **54**(6): 911-929.

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

Basra, A. S. and S. Saha (1999). Growth Regulation of Cotton Fibres. Cotton Fibres, Developmental Biology, Quality Improvement and Textile Processing. A. S. Basra. New York, The Hawthorn Press, Inc.: 47-58.

Baudry, A., M. A. Heim, B. Dubreucq, M. Caboche, B. Weisshaar and L. Lepiniec (2004). "TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis* thaliana." Plant J **39**(3): 366-80.

Beasley, C. A. (1975). "Developmental morphology of cotton flowers and seeds as seen with the scanning electron microscope." American Journal of Botany **62**: 584-592.

Beasley, C. A. and I. P. Ting (1973). "The effects of plant growth substances on *in vitro* fibre development from fertilised cotton ovules." Amer J Bot **60**: 130-139.

Beasley, C. A. and I. P. Ting (1974). "Effects of plant growth substances on *in vitro* fibre development from unfertilised cotton ovules." Amer J Bot **61**: 188-194.

Berlin, J. (1986). The outer epidermis of the cottonseed. Cotton Physiology. The Cotton Foundation. S. J. M. Mauney JR. Memphis: 375–414.

Berliner, E. (1915). "Über die schlaffsucht der mehlmottenraupe (Ephestia kuhniella, Zell.) und ihren erreger Bacillus thuringiensis n. sp." Z. Agnew. Entomol.(2): 29–35.

Bharti, A. K. and J. P. Khurana (2003). "Molecular characterization of transparent testa (tt) mutants of *Arabidopsis* thaliana (ecotype Estland) impaired in flavonoid biosynthetic pathway." Plant Science **165**(6): 1321-1332.

Bundock, P., A. Dendulkras, A. Beijersbergen and P. J. J. Hooykaas (1995). "Transkingdom T-DNA Transfer from Agrobacterium-Tumefaciens to Saccharomyces-Cerevisiae." Embo Journal **14**(13): 3206-3214.

Burn, J. E., U. A. Hurley, R. J. Birch, T. Arioli, A. Cork and R. E. Williamson (2002). "The cellulose-deficient *Arabidopsis* mutant rsw3 is defective in a gene encoding a putative glucosidase II, an enzyme processing N-glycans during ER quality control." Plant Journal **32**(6): 949-960.

Carey, C. C., J. T. Strahle, D. A. Selinger and V. L. Chandler (2004). "Mutations in the pale aleurone color1 regulatory gene of the Zea mays anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in *Arabidopsis* thaliana." Plant Cell **16**(2): 450-464.

Carpenter, R. and E. S. Coen (1990). "Floral Homeotic Mutations Produced by Transposon-Mutagenesis in Antirrhinum-Majus." Genes & Development **4**(9): 1483-1493.

Casas-Mollano, A. and L. Destefano-Beltrán (2000). "Characterization of a cDNA Encoding a TTG1-like Protein (Accesion No. AF220203) from Malus domestica Fruits." Plant Physiol **122**: 1458.

Chalmers, P. (2004). A study of TTG1 homologues in cotton. School of Molecular and Biomedical Science. Adelaide, University of Adelaide.

Chevalier, J., J. Yi, O. Michel and X. M. Tang (1997). "Biotin and digoxigenin as labels for light and electron microscopy in situ hybridization probes: where do we stand?" J Histochem Cytochem **45**(4): 481-91.

Clough, S. J. and A. F. Bent (1998). "Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis* thaliana." Plant Journal **16**(6): 735-743.

Corley, S. B., R. Carpenter, L. Copsey and E. Coen (2005). "Floral asymmetry involves an interplay between TO and MYB transcription factors in Antirrhinum." Proceedings of the National Academy of Sciences of the United States of America **102**(14): 5068-5073.

Cox, R. A. (1968). "The use of guanidinium chloride in the isolation of nucleic acids." Methods Enzymol **12B**: 120-129.

Crawford, B. C. W., U. Nath, R. Carpenter and E. S. Coen (2004). "CINCINNATA controls both cell differentiation and growth in petal lobes and leaves of antirrhinum." Plant Physiology **135**(1): 244-253.

Cronn, R. C., R. L. Small and J. F. Wendel (1999). "Duplicated genes evolve independently after polyploid formation in cotton." Proc Natl Acad Sci U S A **96**(25): 14406-11.

Cubas, P., N. Lauter, J. Doebley and E. Coen (1999). "The TCP domain: a motif found in proteins regulating plant growth and development." Plant Journal **18**(2): 215-222.

Dagerlind, A., K. Friberg, A. J. Bean and T. Hokfelt (1992). "Sensitive mRNA detection using unfixed tissue: combined radioactive and non-radioactive in situ hybridization histochemistry." Histochemistry **98**(1): 39-49.

de Groot, M. J. A., P. Bundock, P. J. J. Hooykaas and A. G. M. Beijersbergen (1998). "Agrobacterium tumefaciens-mediated transformation of filamentous fungi." Nature Biotechnology **16**(9): 839-842.

Debeaujon, I., K. M. Leon-Kloosterziel and M. Koornneef (2000). "Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*." Plant Physiology **122**(2): 403-413.

Debeaujon, I., A. J. M. Peeters, K. M. Leon-Kloosterziel and M. Koornneef (2001). "The TRANSPARENT TESTA12 gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium." Plant Cell **13**(4): 853-871.

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

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

Delmer, D. P. (1999). Cellulose biosynthesis in developing cotton fibres. Cotton fibers: developmental biology, quality improvement and textile processing. A. S. Basra. New York, Haworth Press.

deVetten, N., F. Quattrocchio, J. Mol and R. Koes (1997). "The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals." Genes & Development **11**(11): 1422-1434.

Deyholos, M. K., G. F. Cavaness, B. Hall, E. King, J. Punwani, J. Van Norman and L. E. Sieburth (2003). "VARICOSE, a WD-domain protein, is required for leaf blade development." Development **130**(26): 6577-6588.

Doebley, J., A. Stec and C. Gustus (1995). "Teosinte Branched1 and the Origin of Maize - Evidence for Epistasis and the Evolution of Dominance." Genetics **141**(1): 333-346.

Dolan, L., C. M. Duckett, C. Grierson, P. Linstead, K. Schneider, E. Lawson, C. Dean, S. Poethig and K. Roberts (1994). "Clonal Relationships and Cell Patterning in the Root Epidermis of *Arabidopsis*." Development **120**(9): 2465-2474.

Duval, M., T. F. Hsieh, S. Y. Kim and T. L. Thomas (2002). "Molecular characterization of AtNAM: a member of the *Arabidopsis* NAC domain superfamily." Plant Molecular Biology **50**(2): 237-248.

Elomaa, P., A. Uimari, M. Mehto, V. A. Albert, R. A. E. Laitinen and T. H. Teeri (2003). "Activation of anthocyanin biosynthesis in Gerbera hybrida (Asteraceae) suggests conserved protein-protein and protein-promoter interactions between the anciently diverged monocots and eudicots." Plant Physiology **133**(4): 1831-1842.

Endrizzi, J. E., E. L. Turcotte and R. J. Kohel (1985). "Genetics, cytology and evolution of *Gossypium*." Advances in Genetics **23**: 271-375.

Epping, B., M. Kittel, B. Ruhnau and V. Hemleben (1990). "Isolation and Sequence-Analysis of a Chalcone Synthase Cdnà of Matthiola-Incana R Br (Brassicaceae)." Plant Molecular Biology **14**(6): 1061-1063.

Esch, J. J., M. Chen, M. Sanders, M. Hillestad, S. Ndkium, B. Idelkope, J. Neizer and M. D. Marks (2003). "A contradictory GLABRA3 allele helps define gene interactions controlling trichome development in *Arabidopsis*." Development **130**(24): 5885-5894.

Eulgem, T., P. J. Rushton, S. Robatzek and I. E. Somssich (2000). "The WRKY superfamily of plant transcription factors." Trends Plant Sci **5**(5): 199-206.

Farquharson, M., R. Harvie and A. M. McNicol (1990). "Detection of messenger RNA using a digoxigenin end labelled oligodeoxynucleotide probe." J Clin Pathol **43**(5): 424-8.

Feinbaum, R. L. and F. M. Ausubel (1988). "Transcriptional Regulation of the *Arabidopsis*-Thaliana Chalcone Synthase Gene." Molecular and Cellular Biology **8**(5): 1985-1992.

Feinberg, A. P. and B. Vogelstein (1983). "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity." Anal Biochem **132**(1): 6-13.

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.

Focks, N., M. Sagasser, B. Weisshaar and C. Benning (1999). "Characterization of tt15, a novel transparent testa mutant of *Arabidopsis* thaliana (L.) Heynh." Planta **208**(3): 352-357.

Fryxell, P. A. (1979). The natural history of the cotton tribe. College Station, TX, Texas A&M University Press.

Fujita, M., Y. Fujita, K. Maruyama, M. Seki, K. Hiratsu, M. Ohme-Takagi, L. S. P. Tran, K. Yamaguchi-Shinozaki and K. Shinozaki (2004). "A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway." Plant Journal **39**(6): 863-876.

Gall, J. G. and M. L. Pardue (1969). "Formation and detection of RNA-DNA hybrid molecules in cytological preparations." Proc Natl Acad Sci U S A **63**(2): 378-83.

Galway, M. E., J. D. Masucci, A. M. Lloyd, V. Walbot, R. W. Davis and J. W. Schiefelbein (1994). "The Ttg Gene Is Required to Specify Epidermal-Cell Fate

and Cell Patterning in the *Arabidopsis* Root." Developmental Biology **166**(2): 740-754.

Gibson, U. E. M., C. A. Heid and P. M. Williams (1996). "A novel method for real time quantitative RT PCR." Genome Research **6**(10): 995-1001.

Giri, A. P. and M. S. Kachole (1998). "Amylase inhibitors of pigeonpea (Cajanus cajan) seeds." Phytochemistry **47**(2): 197-202.

Gleave, A. P. (1992). "A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome." Plant Mol Biol **20**(6): 1203-7.

Glover, B. J., M. Perez-Rodriguez and C. Martin (1998). "Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor." Development **125**(17): 3497-508.

Goff, S. A., K. C. Cone and V. L. Chandler (1992). "Functional-Analysis of the Transcriptional Activator Encoded by the Maize-B Gene - Evidence for a Direct Functional Interaction between 2 Classes of Regulatory Proteins." Genes & Development **6**(5): 864-875.

Grbavec, D. and S. Stifani (1996). "Molecular interaction between TLE1 and the carboxyl-terminal domain of HES-1 containing the WRPW motif." Biochemical and Biophysical Research Communications **223**(3): 701-705.

Greve, K., T. La Cour, M. K. Jensen, F. M. Poulsen and K. Skriver (2003). "Interactions between plant RING-H2 and plant-specific NAC (NAM/ATAF1/2/CUC2) proteins: RING-H2 molecular specificity and cellular localization." Biochemical Journal **371**: 97-108.

Grotewold, E., M. B. Sainz, L. Tagliani, J. M. Hernandez, B. Bowen and V. L. Chandler (2000). "Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R." Proceedings of the National Academy of Sciences of the United States of America **97**(25): 13579-13584.

Gruber, M. Y., S. Wang, S. Ethier, J. Holowachuk, P. C. Bonham-Smith, J. Soroka and A. Lloyd (2006). ""HAIRY CANOLA" - *Arabidopsis* GL3 induces a dense covering of trichomes on Brassica napus seedlings." Plant Molecular Biology **60**(5): 679-698.

Grula, J. W., R. L. Hudspeth, S. L. Hobbs and D. M. Anderson (1995). "Organization, Inheritance and Expression of Acetohydroxyacid Synthase Genes in the Cotton Allotetraploid Gossypium-Hirsutum." Plant Molecular Biology **28**(5): 837-846.

Haigler, C. H., D. H. Zhang and C. G. Wilkerson (2005). "Biotechnological improvement of cotton fibre maturity." Physiologia Plantarum **124**(3): 285-294.

Halcomb, J. L., J. H. Benedict, B. Cook and D. R. Ring (1996). "Survival and growth of bollworm and tobacco budworm on nontransgenic and transgenic cotton expressing a CryIA insecticidal protein (Lepidoptera: Noctuidae)." Environmental Entomology **25**(2): 250-255.

Hanahan, D. (1983). "Studies on transformation of Escherichia coli with plasmids." J Mol Biol **166**(4): 557-80.

Hanke, D. E. and D. H. Northcote (1975). "Molecular Visualization of Pectin and DNA by Ruthenium Red." Biopolymers **14**(1): 1-17.

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

Harpster, M. H., J. A. Townsend, J. D. G. Jones, J. Bedbrook and P. Dunsmuir (1988). "Relative Strengths of the S-35 Cauliflower Mosaic-Virus, 1', 2', and Nopaline Synthase Promoters in Transformed Tobacco, Sugarbeet and Oilseed Rape Callus-Tissue." Molecular & General Genetics **212**(1): 182-190.

He, C. X., J. Q. Yan, G. X. Shen, L. H. Fu, A. S. Holaday, D. Auld, E. Blumwald and H. Zhang (2005a). "Expression of an *Arabidopsis* vacuolar sodium/proton antiporter gene in cotton improves photosynthetic performance under salt conditions and increases fiber yield in the field." Plant and Cell Physiology **46**(11): 1848-1854.

He, X. J., R. L. Mu, W. H. Cao, Z. G. Zhang, J. S. Zhang and S. Y. Chen (2005b). "AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development." Plant Journal **44**(6): 903-916.

Heid, C. A., J. Stevens, K. J. Livak and P. M. Williams (1996). "Real time quantitative PCR." Genome Research **6**(10): 986-994.

Heim, M. A., M. Jakoby, M. Werber, C. Martin, B. Weisshaar and P. C. Bailey (2003). "The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity." Molecular Biology and Evolution **20**(5): 735-747.

Hernandez, J., M. Pizzirusso and E. Grotewold (2000). "The maize Mp1 gene encodes a WD-repeat protein similar to An11 and TTG." Maize Genetics Cooperation Newsletter **74**: 24.

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

Hoffman, C. S. and F. Winston (1987). "A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli." Gene **57**(2-3): 267-72.

Horwitz, J. P., J. Chua, M. Noel, J. T. Donatti and J. Freisler (1966). "Substrates for cytochemical demonstration of enzyme activity. II. Some dihalo-3-indolyl phosphates and sulfates." J Med Chem **9**(3): 447.

Houben, A., S. J. Orford and J. N. Timmis (2005). In Situ Hybridization to Plant Tissues and Chromosomes. In Situ Hybridization Protocols. Melbourne, Humana Press. **326:** 203-218.

Hsieh, T. F., O. Hakim, N. Ohad and R. L. Fischer (2003). "From flour to flower: how Polycomb group proteins influence multiple aspects of plant development." Trends in Plant Science **8**(9): 439-445.

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**(1): 63-70.

Hsu, C. Y., J. N. Jenkins, S. Saha and D. P. Ma (2005). "Transcriptional regulation of the lipid transfer protein gene LTP3 in cotton fibers by a novel MYB protein." Plant Science **168**(1): 167-181.

Hülskamp, M., S. Miséra and G. Jürgens (1994). "Genetic dissection of trichome cell development in *Arabidopsis*." Cell **76**: 555-566.

Humphries, J. A. (2001). Funtional analysis of TTG1 homologues in cotton, University of Adelaide.

Humphries, J. A., A. R. Walker, J. N. Timmis and S. J. Orford (2005). "Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis* thaliana TRANSPARENT TESTA GLABRA1 (TTG1) gene." Plant Mol Biol **57**(1): 67-81.

Inoue, H., H. Nojima and H. Okayama (1990). "High efficiency transformation of Escherichia coli with plasmids." Gene **96**(1): 23-8.

James, P., J. Halladay and E. A. Craig (1996). "Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast." Genetics **144**(4): 1425-1436.

Jang, I. C., J. Y. Yang, H. S. Seo and N. H. Chua (2005). "HFR1 is targeted by COP1 E3 ligase for post-translational proteolysis during phytochrome A signaling." Genes & Development **19**(5): 593-602.

Jaradat, T. and R. Allen (1999). "Isolation of a Novel cDNA Encoding an Auxin-Induced Basic Helix-Loop-Helix Transcription Factor (Accession No. AF165924) from Cotton (Gossypium hirsutum L.)." Plant Physiol **121**: 685.

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**(19): 10614-10619.

Ji, S. J., Y. C. Lu, J. X. Feng, G. Wei, J. Li, Y. H. Shi, Q. Fu, D. Liu, J. C. Luo and Y. X. Zhu (2003). "Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array." Nucleic Acids Res **31**(10): 2534-2543.

Jiang, C. X., R. J. Wright, K. M. El-Zik and A. H. Paterson (1998). "Polyploid formation created unique avenues for response to selection in Gossypium (cotton)." Proceedings of the National Academy of Sciences of the United States of America **95**(8): 4419-4424.

John, H. A., M. L. Birnstiel and K. W. Jones (1969). "RNA-DNA hybrids at the cytological level." Nature **223**(206): 582-7.

John, M. E. and G. Keller (1995). "Characterization of mRNA for a proline-rich protein of cotton fiber." Plant Physiol **108**(2): 669-76.

Kappert, J. (1949). "Die Genetik des incana-Charakters und der Anthocyanbildung bei der Levkoje." Züchter **19**: 289-297.

Kikuchi, K., M. Ueguchi-Tanaka, K. T. Yoshida, Y. Nagato, M. Matsusoka and H. Y. Hirano (2000). "Molecular analysis of the NAC gene family in rice." Molecular and General Genetics **262**(6): 1047-1051.

Kim, H. J. and B. A. Triplett (2004). "Cotton fiber germin-like protein. I. Molecular cloning and gene expression." Planta **218**(4): 516-524.

Kirik, V., M. M. Lee, K. Wester, U. Herrmann, Z. G. Zheng, D. Oppenheimer, J. Schiefelbein and M. Hulskamp (2005). "Functional diversification of MYB23 and GL1 genes in trichome morphogenesis and initiation." Development **132**(7): 1477-1485.

Kirik, V., M. Simon, M. Huelskamp and J. Schiefelbein (2004a). "The ENHANCER OF TRY AND CPCl gene acts redundantly with TRIPTYCHON and CAPRICE in trichome and root hair cell patterning in *Arabidopsis*." Developmental Biology **268**(2): 506-513.

Kirik, V., M. Simon, K. Wester, J. Schiefelbein and M. Hulskamp (2004b). "ENHANCER of TRY and CPC 2 (ETC2) reveals redundancy in the region-specific control of trichome development of *Arabidopsis*." Plant Molecular Biology **55**(3): 389-398.

Kitamura, S., N. Shikazono and A. Tanaka (2004). "TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*." Plant Journal **37**(1): 104-114.

Kohel, R. J. (1999). Cotton germplasm resources and the potential for improved fibre productivity and quality. Cotton fibers:developmental biology, quality improvement and textile processing. A. S. Basra. New York, Haworth Press.

Koornneef, M. (1981). "The complex syndrome of TTG mutants." Arab Inf Serv **18**: 45–51.

Koornneef, M., L. W. M. Dellaert and J. H. Vanderveen (1982). "Ems-Induced and Radiation-Induced Mutation Frequencies at Individual Loci in *Arabidopsis*-Thaliana (L) Heynh." Mutation Research **93**(1): 109-123.

Kosugi, S. and Y. Ohashi (1997). "PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene." Plant Cell **9**(9): 1607-1619.

Kunik, T., T. Tzfira, Y. Kapulnik, Y. Gafni, C. Dingwall and V. Citovsky (2001). "Genetic transformation of HeLa cells by Agrobacterium." Proceedings of the National Academy of Sciences of the United States of America **98**(4): 1871-1876.

Lacroix, B., T. Tzfira, A. Vainstein and V. Citovsky (2006). "A case of promiscuity: Agrobacterium's endless hunt for new partners." Trends in Genetics **22**(1): 29-37.

Lam, E. (2004). "Controlled cell death, plant survival and development." Nat Rev Mol Cell Biol **5**(4): 305-15.

Larkin, J. C., M. D. Marks, J. Nadeau and F. Sack (1997). "Epidermal cell fate and patterning in leaves." Plant Cell **9**(7): 1109-1120.

Larkin, J. C., D. G. Oppenheimer, A. M. Lloyd, E. T. Paparozzi and M. D. Marks (1994). "Roles of the Glabrous1 and Transparent Testa Glabra Genes in *Arabidopsis* Trichome Development." Plant Cell **6**(8): 1065-1076.

Larkin, J. C., D. G. Oppenheimer, S. Pollock and M. D. Marks (1993). "*Arabidopsis GLABROUS1* gene requires downstream sequences for function." Plant Cell **5**: 1739-1748.

Larkin, J. C., J. D. Walker, A. C. Bolognesi-Winfield, J. C. Gray and A. R. Walker (1999). "Allele-specific interactions between ttg and gl1 during trichome development in *Arabidopsis* thaliana." Genetics **151**(4): 1591-604.

Lawrence, S. (2005). "Agbio keeps on growing." Nat Biotechnol **23**(3): 281.

Lazo, G. R., P. A. Stein and R. A. Ludwig (1991). "A DNA transformation-competent *Arabidopsis* genomic library in Agrobacterium." Biotechnology (N Y) **9**(10): 963-7.

Leemans, J., R. Deblaere, L. Willmitzer, H. Degreve, J. P. Hernalsteens, M. Vanmontagu and J. Schell (1982). "Genetic Identification of Functions of Tl-DNA Transcripts in Octopine Crown Galls." Embo Journal **1**(1): 147-152.

Li, B. and S. Fields (1993). "Identification of Mutations in P53 That Affect Its Binding to Sv40 Large T-Antigen by Using the Yeast 2-Hybrid System." Faseb Journal **7**(10): 957-963.

Li, C. H., Y. Q. Zhu, Y. L. Meng, J. W. Wang, K. X. Xu, T. Z. Zhang and X. Y. Chen (2002a). "Isolation of genes preferentially expressed in cotton fibers by cDNA filter arrays and RT-PCR." Plant Science **163**(6): 1113-1120.

Li, D. and R. Roberts (2001). "WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases." Cellular and Molecular Life Sciences **58**(14): 2085-2097.

Li, X., X. D. Wang, X. Zhao and Y. Dutt (2004). "Improvement of cotton fiber quality by transforming the acsA and acsB genes into Gossypium hirsutum L. by means of vacuum infiltration." Plant Cell Rep **22**(9): 691-7. Epub 2004 Jan 23.

Li, X. B., L. Cai, N. H. Cheng and J. W. Liu (2002b). "Molecular Characterization of the Cotton GhTUB1 Gene That Is Preferentially Expressed in Fiber." Plant Physiol **130**(2): 666-74.

Lin, Z., D. He, X. Zhang, Y. Nie, X. Guo, C. Feng and J. M. Stewart (2005). "Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD." Plant Breeding **124**(2): 180-187.

Littlewood, T. and G. Evan (1998). Helix-loop-helix transcription factors. London, Oxford University Press.

Liu, H. C., R. G. Creech, J. N. Jenkins and D. P. Ma (2000). "Cloning and promoter analysis of the cotton lipid transfer protein gene Ltp3." Biochim Biophys Acta **1487**(1): 106-11.

Lloyd, A. M., V. Walbot and R. W. Davis (1992). "*Arabidopsis* and Nicotiana Anthocyanin Production Activated by Maize Regulator-R and Regulator-C1." Science **258**(5089): 1773-1775.

Loguerico, L. L., J. Q. Zhang and T. A. Wilkins (1999). "Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (Gossypium hirsutum L.)." Mol Gen Genet **261**(4-5): 660-71.

Ludwig, S. R. and S. R. Wessler (1990). "Maize R-Gene Family - Tissue-Specific Helix-Loop-Helix Proteins." Cell **62**(5): 849-851.

Ma, D. P., H. Tan, Y. Si, R. G. Creech and J. N. Jenkins (1995). "Differential expression of a lipid transfer protein gene in cotton fibre." Biochim Biophys Acta **1257**: 81-84.

Marks, M. D. and K. A. Feldmann (1989). "Trichome Development in *Arabidopsis* Thaliana .1. T-DNA Tagging of the Glabrous1 Gene." Plant Cell **1**(11): 1043-1050.

Marrison, J. L. and R. M. Leech (1994). "The Subcellular and Intra-Organelle Recognition of Nuclear and Chloroplast Transcripts in Developing Leaf-Cells." Plant Journal **6**(4): 605-614.

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

Mekala, K., A. Knutson and C. Smith (2003). Screening for resistance to cotton leafhopper. Proceedings of the 2003 Beltwide Cotton Production Conference.

Morita, Y., M. Saitoh, A. Hoshino, E. Nitasaka and S. Iida (2006). "Isolation of cDNAs for R2R3-MYB, bHLH and WDR transcriptional regulators and identification of c and ca mutations conferring white flowers in the Japanese morning glory." Plant and Cell Physiology **47**(4): 457-470.

Morris, R. G., M. J. Arends, P. E. Bishop, K. Sizer, E. Duvall and C. C. Bird (1990). "Sensitivity of digoxigenin and biotin labelled probes for detection of human papillomavirus by in situ hybridisation." J Clin Pathol **43**(10): 800-5.

Neer, E. J., C. J. Schmidt, R. Nambudripad and T. F. Smith (1994). "The Ancient Regulatory-Protein Family of Wd-Repeat Proteins." Nature **371**(6495): 297-300.

Nesi, N., I. Debeaujon, C. Jond, G. Pelletier, M. Caboche and L. Lepiniec (2000). "The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in *Arabidopsis* siliques." Plant Cell **12**(10): 1863-1878.

Nesi, N., I. Debeaujon, C. Jond, A. J. Stewart, G. I. Jenkins, M. Caboche and L. Lepiniec (2002). "The TRANSPARENT TESTA16 locus encodes the *ARABIDOPSIS* BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat." Plant Cell **14**(10): 2463-2479.

Nesi, N., C. Jond, I. Debeaujon, M. Caboche and L. Lepiniec (2001). "The *Arabidopsis* TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed." Plant Cell **13**(9): 2099-2114.

Oh, S. K., S. Lee, S. Yu and D. Choi (2005). "Expression of a novel NAC domain-containing transcription factor (CaNAC1) is preferentially associated with incompatible interactions between chili pepper and pathogens." Planta **222**(5): 876-887.

Olsen, A. N., H. A. Ernst, L. Lo Leggio and K. Skriver (2005). "NAC transcription factors: structurally distinct, functionally diverse." Trends in Plant Science **10**(2): 79-87.

Ooka, H., K. Satoh, K. Doi, T. Nagata, Y. Otomo, K. Murakami, K. Matsubara, N. Osato, J. Kawai, P. Carninci, Y. Hayashizaki, K. Suzuki, K. Kojima, Y. Takahara, K. Yamamoto and S. Kikuchi (2003). "Comprehensive analysis of NAC family genes in Oryza sativa and *Arabidopsis* thaliana." DNA Research **10**(6): 239-247.

Oppenheimer, D. G., J. C. Larkin, S. M. Pollock and M. D. Marks (1992). "Regulation of Trichome Differentiation in *Arabidopsis*." Molecular Biology of the Cell **3**: A99-A99.

Orford, S. J. and J. N. Timmis (2000). "Expression of a lipid transfer protein gene family during cotton fibre development." Biochim Biophys Acta **1483**(2): 275-84.

Paroush, Z., R. L. Finley, T. Kidd, S. M. Wainwright, P. W. Ingham, R. Brent and D. Ishhorowicz (1994). "Groucho Is Required for Drosophila Neurogenesis, Segmentation, and Sex Determination and Interacts Directly with Hairy-Related Bhlh Proteins." Cell **79**(5): 805-815.

Paterson, A. H., Y. Saranga, M. Menz, C. X. Jiang and R. J. Wright (2003). "QTL analysis of genotype x environment interactions affecting cotton fiber quality." Theoretical and Applied Genetics **106**(3): 384-396.

Payne, C. T., F. Zhang and A. M. Lloyd (2000). "GL3 encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with GL1 and TTG1." Genetics **156**(3): 1349-1362.

Payne, T., J. Clement, D. Arnold and A. Lloyd (1999). "Heterologous myb genes distinct from GL1 enhance trichome production when overexpressed in Nicotiana tabacum." Development **126**(4): 671-82.

Pear, J. R., Y. Kawagoe, W. E. Schreckengost, D. P. Delmer and D. M. Stalker (1996). "Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase." Proc Natl Acad Sci U S A **93**(22): 12637-42.

Pelletier, M. K. and B. W. Shirley (1996). "Analysis of flavanone 3-hydroxylase in *Arabidopsis* seedlings - Coordinate regulation with chalcone synthase and chalcone isomerase." Plant Physiology **111**(1): 339-345.

Penfield, S., R. C. Meissner, D. A. Shoue, N. C. Carpita and M. W. Bevan (2001). "MYB61 Is Required for Mucilage Deposition and Extrusion in the *Arabidopsis* Seed Coat." Plant Cell **13**(12): 2777-91.

Perez-Rodriguez, M., F. W. Jaffe, E. Butelli, B. J. Glover and C. Martin (2005). "Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of Antirrhinum majus flowers." Development **132**(2): 359-370.

Pourcel, L., J. M. Routaboul, L. Kerhoas, M. Caboche, L. Lepiniec and I. Debeaujon (2005). "TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat." Plant Cell **17**(11): 2966-2980.

Quattrocchio, F., W. Verweij, A. Kroon, C. Spelt, J. Mol and R. Koes (2006). "PH4 of petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway." Plant Cell **18**(5): 1274-1291.

Quattrocchio, F., J. Wing, K. van der Woude, E. Souer, N. de Vetten, J. Mol and R. Koes (1999). "Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color." Plant Cell **11**(8): 1433-1444.

Ramsay, N. A. and B. J. Glover (2005). "MYB-bHLH-WD40 protein complex and the evolution of cellular diversity." Trends in Plant Science **10**(2): 63-70.

Ramsay, N. A., A. R. Walker, M. Mooney and J. C. Gray (2003). "Two basic-helix-loop-helix genes (MYC-146 and GL3) from *Arabidopsis* can activate anthocyanin biosynthesis in a white-flowered Matthiola incana mutant." Plant Molecular Biology **52**(3): 679-688.

Reeves, P. A. and R. G. Olmstead (2003). "Evolution of the TCP gene family in asteridae: Cladistic and network approaches to understanding regulatory gene family diversification and its impact on morphological evolution." Molecular Biology and Evolution **20**(12): 1997-2009.

Rerie, W. G., K. A. Feldmann and M. D. Marks (1994). "The Glabra2 Gene Encodes a Homeo Domain Protein Required for Normal Trichome Development in *Arabidopsis*." Genes & Development **8**(12): 1388-1399.

Riechmann, J. L., J. Heard, G. Martin, L. Reuber, C. Z. Jiang, J. Keddie, L. Adam, O. Pineda, O. J. Ratcliffe, R. R. Samaha, R. Creelman, M. Pilgrim, P. Broun, J. Z. Zhang, D. Ghandehari, B. K. Sherman and C. L. Yu (2000). "*Arabidopsis*

transcription factors: Genome-wide comparative analysis among eukaryotes." Science **290**(5499): 2105-2110.

Robertson, M. (2004). "Two transcription factors are negative regulators of gibberellin response in the HvSPY-signaling pathway in barley aleurone." Plant Physiology **136**(1): 2747-2761.

Rong, J. K., C. Abbey, J. E. Bowers, C. L. Brubaker, C. Chang, P. W. Chee, T. A. Delmonte, X. L. Ding, J. J. Garza, B. S. Marler, C. H. Park, G. J. Pierce, K. M. Rainey, V. K. Rastogi, S. R. Schulze, N. L. Trolinder, J. F. Wendel, T. A. Wilkins, T. D. Williams-Coplin, R. A. Wing, R. J. Wright, X. P. Zhao, L. H. Zhu and A. H. Paterson (2004). "A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (Gossypium)." Genetics **166**(1): 389-417.

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**(2): 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**(1): 47-60.

Ruan, Y. L., D. J. Llewellyn and R. T. Furbank (2003). "Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development." Plant Cell **15**(4): 952-64.

Sagasser, M., G. H. Lu, K. Hahlbrock and B. Weisshaar (2002). "A-thaliana TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins." Genes & Development **16**(1): 138-149.

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

Schellmann, S. and M. Hulskamp (2005). "Epidermal differentiation: trichomes in *Arabidopsis* as a model system." International Journal of Developmental Biology **49**(5-6): 579-584.

Schellmann, S., A. Schnittger, V. Kirik, T. Wada, K. Okada, A. Beermann, J. Thumfahrt, G. Jurgens and M. Hulskamp (2002). "TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*." Embo Journal **21**(19): 5036-5046.

Schiefelbein, J. and M. M. Lee (2006). "A novel regulatory circuit specifies cell fate in the *Arabidopsis* root epidermis." Physiologia Plantarum **126**(4): 503-510.

Schmidhauser, T. J. and D. R. Helinski (1985). "Regions of Broad-Host-Range Plasmid Rk2 Involved in Replication and Stable Maintenance in 9 Species of Gram-Negative Bacteria." Journal of Bacteriology **164**(1): 446-455.

Schnittger, A., U. Folkers, B. Schwab, G. Jurgens and M. Hulskamp (1999). "Generation of a spacing pattern: The role of TRIPTYCHON in trichome patterning in *Arabidopsis*." Plant Cell **11**(6): 1105-1116.

Senchina, D. S., I. Alvarez, R. C. Cronn, B. Liu, J. Rong, R. D. Noyes, A. H. Paterson, R. A. Wing, T. A. Wilkins and J. F. Wendel (2003). "Rate variation among nuclear genes and the age of polyploidy in Gossypium." Molecular Biology and Evolution **20**: 633-643.

Serebriiskii, I., J. Estojak, M. Berman and E. A. Golemis (2000). "Approaches to detecting false positives in yeast two-hybrid systems." Biotechniques **28**(2): 328-+.

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**(2).

Shi, D. Q., J. Liu, Y. H. Xiang, D. Ye, V. Sundaresan and W. C. Yang (2005). "SLOW WALKER1, essential for gametogenesis in *Arabidopsis*, encodes a WD40 protein involved in 18S ribosomal RNA biogenesis." Plant Cell **17**(8): 2340-2354.

Shikazono, N., Y. Yokota, S. Kitamura, C. Suzuki, H. Watanabe, S. Tano and A. Tanaka (2003). "Mutation rate and novel tt mutants of *Arabidopsis* thaliana induced by carbon." Genetics **163**(4): 1449-1455.

Shimizu, Y., S. Aotsuka, O. Hasegawa, T. Kawada, T. Sakuno, F. Sakai and T. Hayashi (1997). "Changes in levels of mRNAs for cell wall-related enzymes in growing cotton fiber cells." Plant and Cell Physiology **38**(3): 375-378.

Shirley, B. W., S. Hanley and H. M. Goodman (1992). "Effects of Ionizing-Radiation on a Plant Genome - Analysis of 2 *Arabidopsis* Transparent-Testa Mutations." Plant Cell **4**(3): 333-347.

Shirley, B. W., W. L. Kubasek, G. Storz, E. Bruggemann, M. Koornneef, F. M. Ausubel and H. M. Goodman (1995). "Analysis of *Arabidopsis* Mutants Deficient in Flavonoid Biosynthesis." Plant Journal **8**(5): 659-671.

Smith, E. F. and C. O. Townsend (1907). "A plant tumor of bacterial origin." Science **25**: 671-673.

Smith, T. F., C. Gaitatzes, K. Saxena and E. J. Neer (1999). "The WD repeat: a common architecture for diverse functions." Trends in Biochemical Sciences **24**(5): 181-185.

Soltis, P. S., D. E. Soltis and M. W. Chase (1999). "Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology.PG - 402-4." Nature **402**(6760): 402-4.

Sompornpailin, K., Y. Makita, M. Yamazaki and K. Saito (2002). "A WD-repeat-containing putative regulatory protein in anthocyanin biosynthesis in Perilla frutescens." Plant Molecular Biology **50**(3): 485-495.

Spelt, C., F. Quattrocchio, J. Mol and R. Koes (2002). "ANTHOCYANIN1 of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms." Plant Cell **14**(9): 2121-2135.

Stewart, J. M. (1975). "Fibre initiation on the cotton ovule (Gossypium hirsutum)." Amer J Bot **62**(7)): 723-730.

Stracke, R., M. Werber and B. Weisshaar (2001). "The R2R3-MYB gene family in *Arabidopsis* thaliana." Current Opinion in Plant Biology **4**(5): 447-456.

Struhl, K. and R. W. Davis (1977). "Production of a functional eukaryotic enzyme in Escherichia coli: cloning and expresion of the yeast structural gene for imidazole-glycerolphosphate dehydratase (his3)." Proc Natl Acad Sci U S A **74**: 5255-5259.

Sugahara, Y., P. Carninci, M. Itoh, K. Shibata, H. Konno, T. Endo, M. Muramatsu and Y. Hayashizaki (2001). "Comparative evaluation of 5 '-end-sequence quality of clones in CAP trapper and other full-length-cDNA libraries." Gene **263**(1-2): 93-102.

Suo, J., X. Liang, L. Pu, Y. Zhang and Y. Xue (2003). "Identification of *GhMYB109* encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (Gossypium hirsutum L.)." Biochim Biophys Acta **1630**: 25-34.

Sweeney, M. T., M. J. Thomson, B. E. Pfeil and S. McCouch (2006). "Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice." Plant Cell **18**(2): 283-294.

Szymanski, D. B., R. A. Jilk, S. M. Pollock and M. D. Marks (1998). "Control of GL2 expression in *Arabidopsis* leaves and trichomes." Development **125**(7): 1161-1171.

Taliercio, E., B. Hendrix and J. M. Stewart (2005). "DNA content and expression of genes related to cell cycling in developing Gossypium hirsutum (Malvaceae) fibers." American Journal of Botany **92**(12): 1942-1947.

Taoka, K., Y. Yanagimoto, Y. Daimon, K. Hibara, M. Aida and M. Tasaka (2004). "The NAC domain mediates functional specificity of CUP-SHAPED COTYLEDON proteins." Plant Journal **40**(4): 462-473.

Teng, S., J. Keurentjes, L. Bentsink, M. Koornneef and S. Smeekens (2005). "Sucrose-specific induction of anthocyanin biosynthesis in *Arabidopsis* requires the MYB75/PAP1 gene." Plant Physiology **139**(4): 1840-1852.

Thomashow, M. F., C. G. Panagopoulos, M. P. Gordon and E. W. Nester (1980). "Host Range of Agrobacterium-Tumefaciens Is Determined by the Ti Plasmid." Nature **283**(5749): 794-796.

Turley, R. B. and R. H. Kloth (2002). "Identification of a third fuzzless seed locus in upland cotton (Gossypium hirsutum L.)." Journal of Heredity **93**(5): 359-364.

Udall, J. A., J. M. Swanson, K. Haller, R. A. Rapp, M. E. Sparks, J. Hatfield, Y. S. Yu, Y. R. Wu, C. Dowd, A. B. Arpat, B. A. Sickler, T. A. Wilkins, J. Y. Guo, X. Y. Chen, J. Scheffler, E. Taliercio, R. Turley, H. McFadden, P. Payton, N. Klueva, R. Allen,

D. S. Zhang, C. Haigler, C. Wilkerson, J. F. Suo, S. R. Schulze, M. L. Pierce, M. Essenberg, H. Kim, D. J. Llewellyn, E. S. Dennis, D. Kudrna, R. Wing, A. H. Paterson, C. Soderlund and J. F. Wendel (2006). "A global assembly of cotton ESTs." Genome Research **16**(3): 441-450.

van Nocker, S. and P. Ludwig (2003). "The WD-repeat protein superfamily in *Arabidopsis*: conservation and divergence in structure and function." Bmc Genomics **4**.

Varma, V. A., C. M. Cerjan, K. L. Abbott and S. B. Hunter (1994). "Non-isotopic in situ hybridization method for mitochondria in oncocytes." J Histochem Cytochem **42**(2): 273-6.

Wada, T., T. Tachibana, Y. Shimura and K. Okada (1997). "Epidermal cell differentiation in *Arabidopsis* determined by a Myb homolog, CPC." Science **277**(5329): 1113-6.

Walker, A. R., P. A. Davison, A. C. Bolognesi-Winfield, C. M. James, N. Srinivasan, T. L. Blundell, J. J. Esch, M. D. Marks and J. C. Gray (1999). "The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein." Plant Cell **11**(7): 1337-1349.

Walker, A. R. and M. D. Marks (2000). "Trichome initiation on *Arabidopsis*." Advances in Botanical Research **31**: 219-236.

Walkerpeach, C. and J. Velten (1994). Agrobacterium-mediated gene transfer to plant cells cointegrate and binary vector systems. Plant Molecular Biology Manual. S. and G. a. R. Schilperoort. Belgium, Kluwer Academic Publishers: 1-19.

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**(1): 7-12.

Wang, M. M. and R. R. Reed (1993). "Molecular-Cloning of the Olfactory Neuronal Transcription Factor Olf-1 by Genetic Selection in Yeast." Nature **364**(6433): 121-126.

Wang, S., J. W. Wang, N. Yu, C. H. Li, B. Luo, J. Y. Gou, L. J. Wang and X. Y. Chen (2004). "Control of plant trichome development by a cotton fiber MYB gene." Plant Cell **16**(9): 2323-34. Epub 2004 Aug 17.

Wendel, J. F. (1989). "New World tetraploid cottons contain Old World cytoplasm." Proc Natl Acad Sci U S A **86**: 4132-4136.

Wendel, J. F., R. C. Cronn, I. Alvarez, B. Liu, R. L. Small and D. S. Senchina (2002). "Intron size and genome size in plants." Molecular Biology and Evolution **19**(12): 2346-2352.

Whittaker, D. J. and B. A. Triplett (1999). "Gene-specific changes in alpha-tubulin transcript accumulation in developing cotton fibers." Plant Physiol **121**(1): 181-8.

Wilkins, T. A. and J. A. Jernstedt (1999). Molecular Genetics of Developing Coton Fibres. Cotton Fibers: Developmental Biology, Quality Improvement and Textile Processing. A. S. Basra. New York, Haworth Press: 231-263.

Wilkins, T. A., K. Rajasekaran and D. M. Anderson (2000). "Cotton Biotechnology." Critical Reviews in Plant Sciences **19**(6): 511-550.

Wilkins, T. A. and C. Y. Wan (1993). "Members of two vacuolar H+-ATPase subunit A gene families are spatially and temporally regulated in developing trichomes." Plant Physiol 102 **(S)**: 10.

Winkel-Shirley, B. (2001). "Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology." Plant Physiology **126**(2): 485-493.

Wolf, P. L., E. Von der Muehll and K. Praisler (1973). "A test for bacterial alkaline phosphatase: use in rapid identification of Serratia organisms." Clin Chem **19**(11): 1248-9.

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**(1): 107-127.

Xie, Q., G. Frugis, D. Colgan and N. H. Chua (2000). "*Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development." Genes & Development **14**(23): 3024-3036.

Yan, J. Q., C. X. He, J. Wang, Z. H. Mao, S. A. Holaday, R. D. Allen and H. Zhang (2004). "Overexpression of the *Arabidopsis* 14-3-3 protein GF14 lambda in cotton leads to a "Stay-Green" phenotype and improves stress tolerance under moderate drought conditions." Plant and Cell Physiology **45**(8): 1007-1014.

Yang, S., F. Cheung, J. Lee, M. Ha, N. Wei, S. Sze, D. Stelly, P. Thaxton, B. Triplett, C. Town and C. Jeffrey (2006). "Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton." Plant Journal **47**(5): 761-775.

Yu, L. H., C. Gaitatzes, E. Neer and T. E. Smith (2000a). "Thirty-plus functional families from a single motif." Protein Science **9**(12): 2470-2476.

Yu, X. H., Y. Q. Zhu, S. Lu, T. Z. Zhang, X. Y. Chen and Z. H. Xu (2000b). "A comparative analysis of a fuzzless-lintless mutant of Gossypium hirsutum L. cv. Xu-142." Science in China Series C-Life Sciences **43**(6): 623-630.

Zhang, D., M. Hrmova, C. H. Wan, C. Wu, J. Balzen, W. Cai, J. Wang, L. D. Densmore, G. B. Fincher, H. Zhang and C. H. Haigler (2004). "Members of a new group of chitinase-like genes are expressed preferentially in cotton cells with secondary walls." Plant Mol Biol **54**(3): 353-72.

Zhang, F., A. Gonzalez, M. Z. Zhao, C. T. Payne and A. Lloyd (2003). "A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*." Development **130**(20): 4859-4869.

Zhang, T. and J. Pan (1991). "Genetic analysis of fuzzless-lintless mutant in upland cotton." Jiangsu J Agric Sci **7**: 14-16.

Zhu, Y. Y., E. M. Machleder, A. Chenchik, R. Li and P. D. Siebert (2001). "Reverse transcriptase template switching: A SMART (TM) approach for full-length cDNA library construction." Biotechniques **30**(4): 892-897.

Zimmermann, I. M., M. A. Heim, B. Weisshaar and J. F. Uhrig (2004). "Comprehensive identification of *Arabidopsis* thaliana MYB transcription factors interacting with R/B-like BHLH proteins." Plant Journal **40**(1): 22-34.

Corrected References

Baudry, A., M. A. Heim, B. Dubreucq, M. Caboche, B. Weisshaar and L. Lepiniec (2004). "TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*." Plant J **39**(3): 366-80.

Berliner, E. (1915). "Über die Schlaffsucht der Mehlmottenraupe (Ephestia kuhniella, Zell.) und ihren Erreger Bacillus thuringiensis n. sp." Z. Angew. Entomol.(2): 29–35.

Bharti, A. K. and J. P. Khurana (2003). "Molecular characterization of transparent testa (tt) mutants of *Arabidopsis thaliana* (ecotype Estland) impaired in flavonoid biosynthetic pathway." Plant Science **165**(6): 1321-1332.

Bundock, P., A. den Dulk-Ras, A. Beijersbergen and P. I. J. Hooykaas (1995). "Transkingdom T-DNA Transfer from Agrobacterium-Tumefaciens to *Saccharomyces cerevisiae*." Embo Journal **14**(13): 3206-3214.

Carey, C. C., J. T. Strahle, D. A. Selinger and V. L. Chandler (2004). "Mutations in the pale aleurone color1 regulatory gene of the Zea mays anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in *Arabidopsis thaliana*." Plant Cell **16**(2): 450-464.

Carpenter, R. and E. S. Coen (1990). "Floral Homeotic Mutations Produced by Transposon-Mutagenesis in *Antirrhinum majus*." Genes & Development **4**(9): 1483-1493.

Clough, S. J. and A. F. Bent (1998). "Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*." Plant Journal **16**(6): 735-743.

Crawford, B. C. W., U. Nath, R. Carpenter and E. S. Coen (2004). "CINCINNATA controls both cell differentiation and growth in petal lobes and leaves of Antirrhinum." Plant Physiology **135**(1): 244-253.

Cronn, R. C., R. L. Small and J. F. Wendel (1999). "Duplicated genes evolve independently after polyploid formation in cotton." Proc Natl Acad Sci U S A **96**(25): 14406-11.

Epping, B., M. Kittel, B. Ruhnau and V. Hemleben (1990). "Isolation and Sequence-Analysis of a Chalcone Synthase cDNA of *Matthiola incana* R Br (Brassicaceae)." Plant Molecular Biology **14**(6): 1061-1063.

Feinbaum, R. L. and F. M. Ausubel (1988). "Transcriptional Regulation of the *Arabidopsis thaliana* Chalcone Synthase Gene." Molecular and Cellular Biology **8**(5): 1985-1992.

Harpster, M. H., J. A. Townsend, J. D. G. Jones, J. Bedbrook and P. Dunsmuir (1988). "Relative Strengths of the S-35 Cauliflower Mosaic-Virus, 1', 2', and Nopaline Synthase Promoters in Transformed Tobacco, Sugarbeet and Oilseed Rape Callus-Tissue." Molecular & General Genetics **212**(1): 182-190.

Humphries, J. A., A. R. Walker, J. N. Timmis and S. J. Orford (2005). "Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana* TRANSPARENT TESTA GLABRA1 (TTG1) gene." Plant Mol Biol **57**(1): 67-81.

Kohel, R. J. (1999). Cotton germplasm resources and the potential for improved fibre productivity and quality. Cotton fibers: developmental biology, quality improvement and textile processing. A. S. Basra. New York, Haworth Press.

Larkin, J. C., J. D. Walker, A. C. Bolognesi-Winfield, J. C. Gray and A. R. Walker (1999). "Allele-specific interactions between ttg and gl1 during trichome development in *Arabidopsis thaliana*." Genetics **151**(4): 1591-604.

Leemans, J., R. Deblaere, L. Willmitzer, H. De Greve, J. P. Hernalsteens, M. Van Montagu and J. Schell (1982). "Genetic Identification of Functions of Tl-DNA Transcripts in Octopine Crown Galls." Embo Journal **1**(1): 147-152.

Lloyd, A. M., V. Walbot and R. W. Davis (1992). "*Arabidopsis* and *Nicotiana* Anthocyanin Production Activated by Maize Regulator-R and Regulator-C1." Science **258**(5089): 1773-1775.

Marks, M. D. and K. A. Feldmann (1989). "Trichome Development in *Arabidopsis thaliana*.1. T-DNA Tagging of the Glabrous1 Gene." Plant Cell **1**(11): 1043-1050.

Mekala, K., A. Knutson and C. Smith (2003). "Screening for resistance to cotton leafhopper." Proceedings of the 2003 Beltwide Cotton Production Conference.

Ooka, H., K. Satoh, K. Doi, T. Nagata, Y. Otomo, K. Murakami, K. Matsubara, N. Osato, J. Kawai, P. Carninci, Y. Hayashizaki, K. Suzuki, K. Kojima, Y. Takahara, K. Yamamoto and S. Kikuchi (2003). "Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*." DNA Research **10**(6): 239-247.

Sagasser, M., G. H. Lu, K. Hahlbrock and B. Weisshaar (2002). "*A. thaliana* TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins." Genes & Development **16**(1): 138-149.

Schellmann, S., A. Schnittger, V. Kirik, T. Wada, K. Okada, A. Beermann, J. Thumfahrt, G. Jurgens and M. Hulskamp (2002). "TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*." EMBO Journal **21**(19): 5036-5046.

Serebriiskii, I., J. Estojak, M. Berman and E. A. Golemis (2000). "Approaches to detecting false positives in yeast two-hybrid systems." Biotechniques **28**(2): 328-232.

Shikazono, N., Y. Yokota, S. Kitamura, C. Suzuki, H. Watanabe, S. Tano and A. Tanaka (2003). "Mutation rate and novel tt mutants of *Arabidopsis thaliana* induced by carbon." Genetics **163**(4): 1449-1455.

Stewart, J. M. (1975). "Fibre initiation on the cotton ovule (*Gossypium hirsutum*)." Amer J Bot **62**(7)): 723-730.

Stracke, R., M. Werber and B. Weisshaar (2001). "The R2R3-MYB gene family in *Arabidopsis thaliana*." Current Opinion in Plant Biology **4**(5): 447-456.

Thomashow, M. F., C. G. Panagopoulos, M. P. Gordon and E. W. Nester (1980). "Host Range of *Agrobacterium tumefaciens* is Determined by the Ti Plasmid." Nature **283**(5749): 794-796.

Walkerpeach, C. and J. Velten (1994). Agrobacterium-mediated gene transfer to plant cells cointegrate and binary vector systems. Plant Molecular Biology Manual. S. and G. a. R. Schilperoort. Belgium, Kluwer Academic Publishers: 1-19.

Wilkins, T. A. and J. A. Jernstedt (1999). Molecular Genetics of Developing Cotton Fibres. Cotton Fibers: Developmental Biology, Quality Improvement and Textile Processing. A. S. Basra. New York, Haworth Press: 231-263.

Zimmermann, I. M., M. A. Heim, B. Weisshaar and J. F. Uhrig (2004). "Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins." Plant Journal **40**(1): 22-34.