Understanding the Molecular Processes That Drive Cotton Fibre Yield and Quality

Danny Llewellyn, Yingru Wu, Adriane Machado, Rosemary White, Yves Al-Ghazi, Todd Collins, Jim Peacock and Elizabeth Dennis
CSIRO plant Industry P.O. Box 1600 Canberra City ACT 2601.

Introduction

The skin or epidermis of a plant is its outward face to the World and buffers it from changes in temperature and humidity; protects it against threats such as pests and pathogens; helps it extract nutrients from the air and soil and even assists it to disperse its seeds into new and hopefully more favourable places to grow. Many plants have evolved specialized epidermal cells (Figure 1) for some of these functions and scientists are starting to unravel at a detailed molecular level the processes that allow the cells in the plant's skin to develop these specialized roles relative to the adjacent cells that are just the bricks and mortar of the plant (Larkin et al., 2003). Cotton is no different and has developed all sorts of specialized hairs that cover the plants leaves, stems, petioles, petals and roots, but it is the very long hairs on the seeds that first attracted man's interest in cotton and these have driven the domestication and widespread adoption of cotton as a premier agricultural crop for fibre production for textiles.

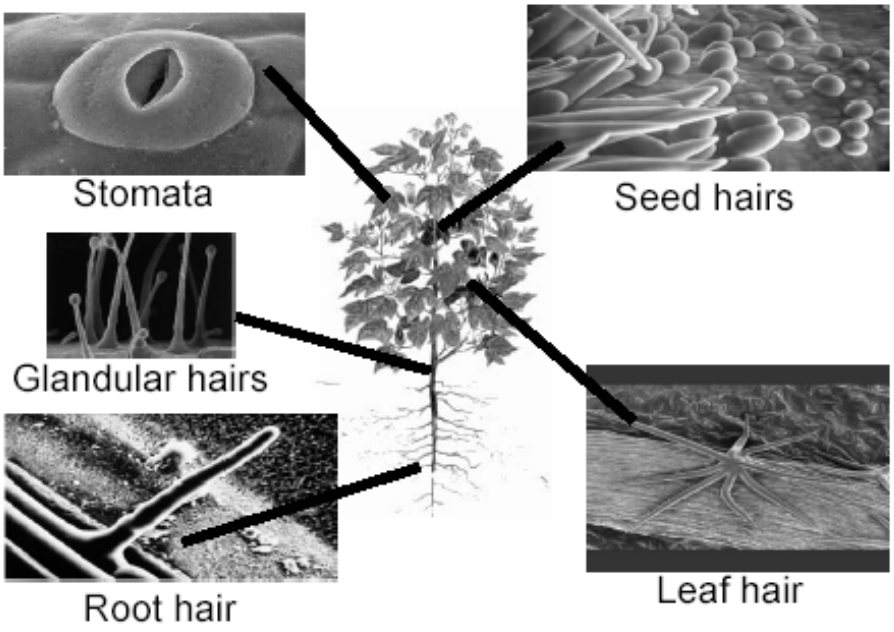

Figure 1. Different sorts of epidermal cells that occur in plants, including cotton. Stomata are the gateways for the exchange of gases into and out of the leaf and other tissues. Leaf hairs and glandular trichomes provide either a physical or chemical defence against insects. Root hairs help extract nutrients and water from the soil. Seed hairs (fibres) are thought to help in the dispersal of seeds in the environment, but are so well developed in cotton that they can be spun into yarn.

Scientists have studied some of these different cell types in model plants like Arabidopsis so we already know quite a bit about how the ordinary epidermal cells are changed into these specialised cells. Unfortunately in cotton fibres we know very little except for very descriptive work about the structures of the fibres as they grow and some of the biochemistry of fibre formation. Plant hormones clearly play an important role in the development of the fibre and this has been studied in detail in tissue culture. If we could understand a lot more about the molecular controls on fibre development we would have a better chance of being able to manipulate fibre development to improve the production and properties of the fibres and so help our cotton breeders to produce higher yielding, higher quality cotton and in the longer term make novel fibres that would differentiate Australian cotton in what has become a very competitive international marketplace.

Cotton Fibre Development

Cotton fibres are single cells that expand out from the surface of the cotton seed on the day that the flower opens (Stewart, 1975) (Figure 2). These cells then extend until they are a few thousand times longer than they are wide (taking about 25 days) and reach the final length of the mature fibre. When this extension process is almost complete the inside of the cell begins to fill with cellulose making the walls of the cells very thick. After another 20 days or so the cells begin to die and dry out to become the flattened tubes of almost pure cellulose that are harvested and spun into yarn for textile production. There is also a second type of fibre, the fuzz fibre, which starts growing a few days after the first lot of fibres, but these short thick cells stop growing after

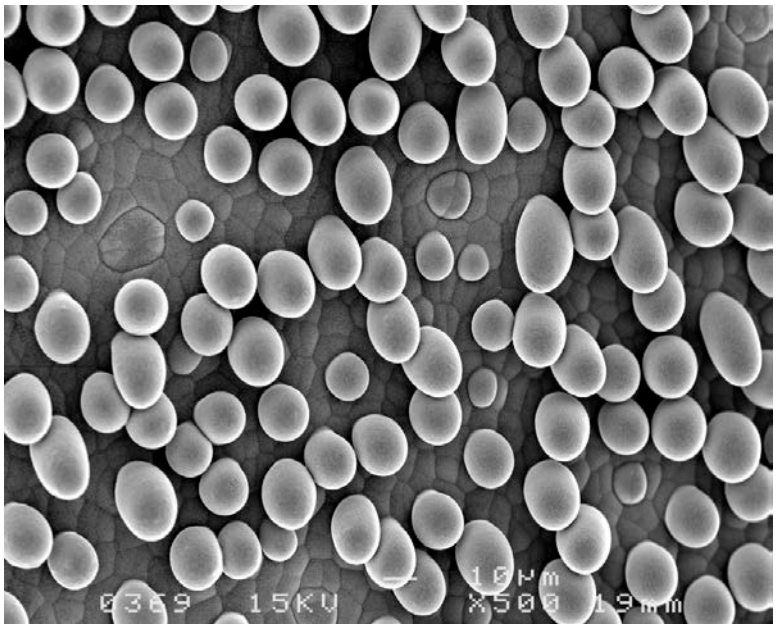

Figure 2. Scanning electron micrograph of the surface of a cotton seed on the day of flower opening. Note the small balloon like cells that have bubbled out from the underlying epidermal cells. These cells then extend longitudinally to become the long fibres harvested from the seeds.

they reach a few millimetres. These cells are the fuzz that is left after the long lint fibre is ginned off the seed. The number of epidermal cells that start to expand, the length of time that the cells expand radially at the start of this process, the time they extend longitudinally and the amount of thick wall they make determine all the important properties of yield of fibre per seed, fineness and maturity (Micronaire), strength, and length that ultimately determine how well you as growers can market your cotton. Our breeding team has done an extremely proficient job in producing new varieties with both higher yields and fibre quality, but as we push for higher and higher yields interactions between different yield components and fibre traits are becoming difficult to balance (there are negative correlations between yield and fibre length, for example)(see Constable et al., 2006: these proceedings). Do we understand anything about the genes that control fibre growth and development and why there are interactions between different traits? No, or at least very little. Some genes involved in fibre production have been discovered, but these are mainly genes for structural proteins involved in the ability of the cells to expand or the enzymes we know must be there to make the proteins and cellulose needed for the growing cell walls, but none of these genes on their own are likely to control the important fibre properties that are of interest to cotton breeders. What we need to get at are the master control genes that determine which cells will initiate the process of fibre development, genes that control the elongation process and genes that control the cell wall thickening process. These are likely to be special classes of genes called transcription factors that produce proteins that bind to the DNA of other genes and cause them to be turned on or off.

There are two ways of getting at the master control genes for fibre yield and quality: either through genetics and molecular (DNA) marker techniques or through genomic or molecular biology techniques. Our breeding program is obviously interested in a marker approach as it will give them useful diagnostic tools for tracking multiple fibre quality traits through their breeding populations. This is still a difficult process in cotton relative to other plants because the low level of genetic diversity and the poor genomic resources for cotton do not allow efficient discovery of fibre quality markers or conversion of DNA marker location information into genes and their functions. Although we have some capabilities in DNA marker research, a significant investment would be required to establish a large marker discovery program in cotton in Australia and for us to catch up with other groups overseas. There is currently a lot of international activity in this area and we may be able through collaboration, at least, to capitalise on that work to help our breeding program in the short term to use markers for fibre quality traits. We hope to establish some new projects in developing useful markers for disease resistance traits in cotton as these could have significant impacts on the efficiency of breeding for resistance to diseases that currently impact on yield.

Developing Genomic Tools for Cotton

The second way of getting at genes that might control fibre growth and fibre quality is the genomic route. CSIRO over the past five years has been establishing a capability in cotton genomics to complement our existing capabilities in cotton biotechnology and breeding. Genomics is the study of plant process at a whole genome level rather than through the detailed study of

single genes and their actions. At its highest level genomics is underpinned by having the whole genome sequence - the sequence of every gene - of an organism. In plants, it is most advanced in the two model species Arabidopsis and rice where there are many resources available, including: the sequence of every gene and its flanking control regions across every chromosome; detailed genetic maps of several different lines or ecotypes; libraries of large fragments of DNA covering the whole genome; libraries of seed lines where every gene has been inactivated by insertion of some foreign DNA; and an assortment of high density DNA chips that allow you to analyse which genes are expressed under any particular treatment condition or developmental stage or cell type (as well as databases to access anything that other researchers have done with the chips or sequences). One need only query a variety of international databases to find out all about any particular gene (for example what tissues it is expressed in or its chromosomal region), order the gene, or seed of a mutant line with that gene inactivated and it can be in your laboratory within days for you to use in your own studies. These resources have come about because of international co-operation and determination and obviously the investment of hundreds of millions of dollars over many years. The same resources are not available for cotton, but some progress is being made with the development of genetic maps, large fragment BAC libraries and collections of the sequences of expressed genes in different tissues (so called ESTs – Expressed Sequence Tags). There is also a proposal to sequence the genome of one of the diploid relatives of cotton (*Gossypium raimondii*) that has recently been funded as a pilot project by the US-DOE before it it considered for full sequencing.

Some benefit can be gained in cotton from using the resources available for Arabidopsis since many genes and their functions will be common between different plants. Specialised structures such as the cotton seed fibres however, are not found in Arabidopsis, so CSIRO has focussed on developing genomic resources that would be of benefit for the study of cotton fibre growth and development. Funding from CSIRO and CRDC has allowed us to develop DNA chips with tens of thousands of fragments of DNA corresponding to the genes that are expressed in the developing cotton seed at the time that the fibres are forming. The sequence of every gene on the CSIRO cotton chip has been determined (CRDC funding only allowed a few thousand sequences to be determined, but by collaborating with scientists at Iowa State and Arizona University who had US National Science Foundation funding we were able to get an additional 14,000 sequences). Our first chip had 10,000 seed genes (Wu et al., 2005), but we have broadened our second generation of chip to contain an additional set of 13,000 genes from elongating fibre purchased from Clemson Genomics Institute as well as genes from leaves, embryos, and seedlings to make a 24,000 genes chip that should also have uses beyond just the study of fibres. The chips are only the size of a microscope slide but now provide us with a tool to simultaneously probe the changing patterns of gene expression of all the genes on the chips (plants probably have about 40,000 genes so we don't yet have complete gene coverage) in particular cells, between different genetic material or in plants treated in different ways.

Gene Chips help identify genes involved in the production of cotton fibres

Our first use for our cotton seed chip was to see if we could identify important genes needed to start the process of fibre development. There are some naturally occurring mutants of

cotton that have almost completely fibreless seeds (Figure 3). If you look closely at these seeds on the day of flowering they have hardly any of the small bubbles of fibre cells and should not express any of the genes needed to make a fibre. DNA chips are ideal for making these comparisons between a mutant and normal line as they identify those genes that are expressed differently between any two conditions under examination. The expressed genes (actually the messenger RNA copy of the gene that is normally translated into the protein product of the gene) from the mutant fibreless seed, for example, are labelled with a red dye and the expressed genes from the seeds of the normal line are labelled with a green dye, mixed and probed onto the chip. Each piece of DNA on the chip (printed as a tiny little dot in a grid) will bind to its corresponding labelled messenger (both the red and green versions). If that particular gene is expressed more in the mutant than in the normal seed then the spot for that gene will have a more reddish tinge or if the if more in the normal than the mutant it will have a greenish colour. Genes that are no different in the mutant and normal seeds will have equal amounts of red and green and appear yellow. So by scanning the slides and analysing the colours of all the spots (each corresponding to a particular gene) we can find which genes are expressed in the normal seed that is making lots of fibres and not in the fibreless mutant – these genes are likely to be the genes that are the ones that are highly expressed in the tiny expanding fibre cells (as this is the only difference between the normal and mutant seeds. Doing the same experiments a few times with different baches of seeds and different mutants (we had six different fibreless mutants) as well as some tricky statistical filtering we were able to come up with a small set of genes that were consistently not expressed in fibreless seeds.

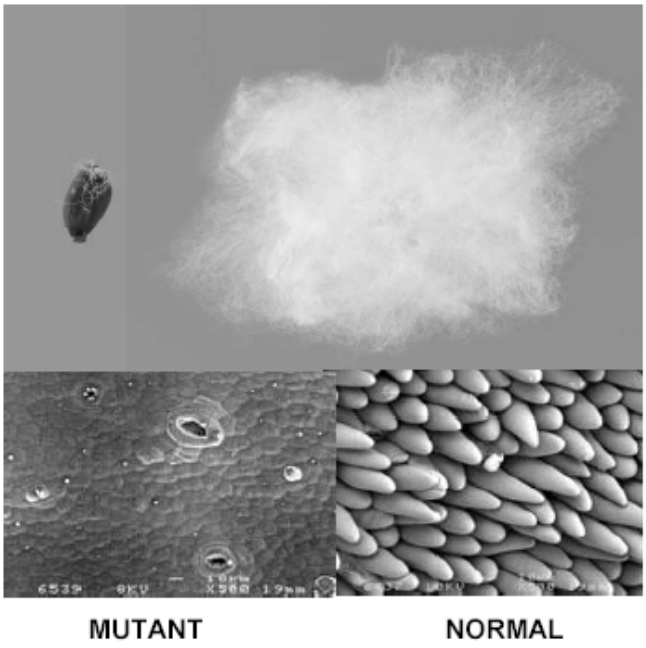

Figure 3. Mature seed of a fibreless mutant of cotton (left) and the normal linted seed (right). Below are the highly magnified images of the surface of the seed late on the day of flowering. The normal seed (right) has already started to produce fibres that are rapidly growing away from the seed. The fibreless mutant on the left has hardly any fibres and the seed surface is flat except for a few stomates (the circular structures).

Out of 10,000 cotton genes only a dozen genes were differentially active between the fibreless and normal linted cotton seeds on the day of flowering when fibres normally start to grow (Wu et al., 2006). What sorts of genes showed a difference in activity? Well, as you might expect some of them were genes that would be expected in a rapidly growing fibre cell – these included expansin (a protein thought to help loosen the cell wall around the cell to allow it to expand from the pressure of the water inside the cell), lipid transfer protein (proteins thought to be needed for the production of more membrane in the expanding cell) and sucrose synthase (an enzyme involved in breaking down sugar needed for generating the pressure inside the cell and for the production of the cellulose for the new cell walls being made as the fibre expands). These types of genes had previously been known to be involved in fibre growth (and both CSIRO and University of Adelaide researchers already had projects working on these genes). The most interesting genes to us, however, were the transcription factor and cell cycle genes that were differently expressed between the mutant and the normal cotton seeds. These are master genes that can affect cascades of other genes and hence might be more amenable to manipulating during breeding or in transgenic cotton plants.

Two different sorts of transcription factors were discovered one type were Myb genes (two different types were discovered) and the second was a homeodomain gene. The two different Myb genes were both significantly under expressed in all the fibreless seeds, and although similar to each other were not identical. Myb genes are a type of transcription factor known to play important roles in all sorts of plant development and biochemical pathways (Arabidopsis has over a hundred different Myb genes – most with functions still unknown, but some regulate cell shape, others leaf and root hair development and still others various secondary chemical defence pathways). The homeodomain genes are also involved in plant development and defence, and perhaps one of the better known genes in plants is one controlling leaf hair development. The similarity of some of these cotton fibre genes with leaf hair (trichome) development is perhaps not very surprising, given what had been known about leaf hair development in Arabidopsis, but these cotton genes are not the same as the Arabidopsis genes and probably have their own unique functions in cotton to make fibres. In addition to these known types of genes we also identified a couple of genes that do not appear to have any directly comparable genes in the international databases of plant and animal genes. These might represent novel genes with as yet undefined functions.

Where to next?

We now have about half a dozen good candidates for genes that might have an important role in determining which cells on the epidermis of a cotton seed will turn into cotton fibres. How do we go about proving they have such roles and start to use that information to assist our breeders? In establishing function, the best way is to use reverse genetics – if you can knock out the function of a particular gene (as happens in naturally or artificially induced mutants) does the loss of that gene have an effect on the plant eg., in our case does it stop the production of the fibres? (Although we see loss of expression of our candidate genes in the fibreless mutants this may not be the cause, but just an effect of fibres not being produced). This strategy has been used to great effect in Arabidopsis as researchers start to unravel what the 30,000 genes in this plant actually do

(even though they have all the sequences we are still unsure of what more than half the genes in Arabidopsis do).

Making a so called knockout or loss of function mutant in a specific gene can be achieved in a number of ways – one is to insert a large piece of foreign DNA into the gene so that its encoded message is disrupted (this has to be done in a shotgun way as foreign DNA is inserted at random into the genome of a transgenic plant and large numbers have to be generated and screened to find one with any particular gene inactivated) – the second is to silence the output of the gene using a different transgenic plant strategy that effectively activates a DNA sequence specific degradation of the message from a particular gene. CSIRO is at the forefront of gene silencing technology and have developed a patented system called RNAi that allows one particular gene sequence at a time to be targeted for silencing. This type of gene silencing has been used to generate the high oleic GM cotton being developed by our colleagues in the CSIRO oilseeds group at Plant Industry. Unfortunately cotton is a very difficult plant to use to individually silence large numbers of genes because it takes a lot of time and effort to make a transgenic cotton plant that carries the silencing construct. Despite this, we have gone ahead and targeted six of our candidates to see what effect silencing has on cotton fibre development.

Preliminary data is starting to come out for our three most interesting transcription factor genes. Silencing of one of the Mybs does cause the production of fibreless seeds, while silencing the other Myb seems to have a number of effects on fibre production (delaying fibre emergence from the seed, reducing overall length), but also has effects on leaf hair and seed production. These other effects may be because there are more than one similar Myb gene in cotton and we may be inadvertently silencing multiple genes in the transgenic RNAi plants. Silencing the homeodomain gene seems to mostly affect leaf hair production, but may also have an effect on seed development. We have to take these plants through a few more generations before we can be confident that the effects observed are caused by the loss of expression of the particular genes and we can then put these transgenic plant lines back through a DNA chip analysis to see what other genes are normally controlled by the particular transcription factors (ie compare the normal seeds with the gene silenced seeds). If less activity of these candidate genes is bad for fibre development then perhaps more of these genes may enhance fibre quality or yield. So, once we can clearly define their roles we can begin to look at what effects changing the expression of those genes might have on fibre production or properties using so called overexpression constructs in a transgenic plant. We can also look at breeding lines and see if the expression of any of these candidate genes correlates with particular fibre properties, including yield and use them as markers for breeding for quality.

One simple set of DNA chip experiments has generated half a dozen good candidate fibre genes to follow up in more detail; with at least one very interesting gene identified that might be a key gene in fibre initiation. Once we start using our latest and larger DNA chip we will probably find more interesting genes and we are also extending our analyses to later stages in fibre development when the fibre cells are rapidly elongating or even later when they are laying down their thick secondary cell walls. Comparing different genotypes that differ considerably in their fibre properties at defined fibre development stages should reveal more interesting genes that play a role in specifying fibre quality traits such as fibre length, fibre strength and fibre fineness and maturity. Integrating such DNA chip studies with more conventional genetics and breeding will

hopefully allow us to identify the most important genes for fibre quality and to assist the breeding program to take the most advantage from existing germplasm to enhance both yield and quality and perhaps even provide a transgenic approach to producing fibres with novel properties.

References

Larkin, J.C., Brown, M.L. and Schiefelbein, J. 2003. How do cells know what they want to be when they grow up? Lessons from epidermal patterning in Arabidopsis. Annu. Rev. Plant Biol. 54: 403-430.

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

Wu, Y., Rozenfeld, S., Defferrard, A., Ruggiero, K., Udall, J. A., Kim, HR., Llewellyn, D.J. and Dennis, E.S. 2005. Cycloheximide treatment of cotton ovules alters the abundance of specific classes of mRNAs and generates novel ESTs for microarray expression profiling. Mol. Genet. Genomics 274: 477-493.

Wu, Y., Machado, A., White, R.G., Llewellyn, D.J. and Dennis, E.S. 2006. Identification of early genes expressed during cotton fibre initiation using cDNA microarrays. Plant Cell Physiology 47: 107-127.