Cotton fiber biology: Integrating genomics and development for improving fiber traits

Thea A. Wilkins
Cotton Functional Genomics Center, Department of Agronomy & Range Science, University of California, Davis CA USA
Correspondence author tawilkins@ucdavis.edu

ABSTRACT

Understanding the underlying genetic mechanisms that control fiber growth and development and govern complex fiber traits is key to implementing long term strategies to improve yield and fiber quality using molecular approaches. As part of our fiber genomics initiative, a gene discovery project generated almost 50,000 ESTs (expressed sequence tags) from rapidly elongating cotton fibers that currently define the cotton fiber transcriptome, which represents ~35% of the cotton genome. The genetic complexity of elongating cotton fibers is very high and reflects the highly exaggerated growth of rapidly elongating fibers, indicating a very metabolically active cell type. Expression profiling of cotton long oligonucleotide arrays representing the fiber transcriptome provides a global view of the thousands of expansion-associated genes that are differentially regulated during fiber development, and provide candidate genes for genetic modification. With a genomic toolbox, candidate genes and a developmental model in hand, we can begin to focus determining fiber gene function using reverse genetic approaches as a step towards genetic modification of the fiber transcriptome. New innovations in cotton transformation and regeneration technology have improved and simplified the introduction of transgenes into cotton, and major strides have been made towards genotype-independent transformation. We are now entering a new era of cotton research in the areas of functional genomics and applied biotechnology with a promising future for manipulating the fiber transcriptome for improvement of fiber traits.

Introduction

The global production of genetically modified cotton is certainly one of biotechnology's greatest success stories (reviewed in Wilkins *et al.*, 2000). In 2002, ~20% of cotton acreage worldwide was planted in transgenic cotton, although the percentage of transgenic cotton in production is significantly higher in some countries. In the U.S. for instance, transgenic cotton accounts for >75% of the acreage. At present, transgenic cotton is genetically modified for input traits, that is, novel traits conferring resistance to herbicides or insects. The widespread acceptance of transgenic cotton by growers is based on increased profit margins through decreasing production costs, the crop is easier to manage, and is generally more environmentally friendly. The success of GMO cotton in the U.S., however, is offset by reports of yield plateaus and declining fiber quality during the same time frame. Genetic erosion has certainly been a major contributing factor, and genetic improvement programs stymied by a narrow germplasm base are scurrying to implement long-term strategies to identify and introgress new sources of diversity into breeding stocks.

The most value derived from the production of cotton is the fiber, and one approach to improving fiber yield and fiber quality that warrants serious consideration is the potential of GMO cotton engineered for output traits to increase productivity (Wilkins *et al.*, 2000). Output traits targeted for improvement include the agronomically important fiber properties such as yield, length, strength, fineness, uniformity, maturity, and brightness, as well as post-harvest properties like dye uptake and retention. Cotton biotechnology also offers opportunities to create "designer" cotton that possesses novel fiber traits that would considerably expand the utilization of cotton fiber. For instance, a fiber that was flame-retardant and exhibited greater elasticity would open new markets to cotton fiber. Other potential "niche" markets include fibers designed specifically for bioremediation or pharmaceutical applications. However, there has been ongoing debate for some time calling into question whether manipulating a simple trait could impact a complex trait like yield and fiber quality. Even so, we believe that a comprehensive understanding of fiber biology, coupled to the judicious selection of candidate genes, poses an opportunity that should be aggressively pursued, especially to complement other breeding approaches in the short term.

One of the main reasons that genetic modification of cotton for output traits has been slow to enter the scene has been the need to build up a Biotechnology Tool Chest. Here we report on research progress that highlights significant advances that bring us to the point that implementing GMO strategies for output traits is no longer a dream, but a reality. The three major topics for discussion are: 1) Understanding fiber biology through gene discovery using genomic approaches, 2) Cotton transformation and regeneration technology, and 3) A pilot study that lays the groundwork for modifying output traits.

Gene discovery defines the cotton fiber transcriptome

Understanding the biology of cotton fibers requires knowing what genes are expressed in fibers, how the genes are regulated both in space and time, and most importantly, what function each gene executes to contribute towards the growth and structure of the fiber and how this correlates to agronomic properties. It has long been speculated that a large number of genes are required to make a cotton fiber, yet to date, fewer than 50 fiber genes are described in the literature using traditional approaches in molecular biology. To

address the needs of the cotton community, we set out to define the cotton fiber transcriptome, that is, all the unique genes that are expressed in the fiber, using genomic approaches. Our goal was to develop a comprehensive database of fiber ESTs (expressed sequence tags) as part of a large-scale gene discovery effort. Because of the redundancy inherent to tetraploid species, we opted to develop cultivated diploid cotton (*Gossypium arboreum* L.) as a model system for cotton functional genomics.

Our approach to maximizing gene discovery to define the cotton fiber transcriptome was multi-pronged and based on the following strategies; 1) Focus on diploid cotton as a model system, 2) Target sequencing of 50,000 ESTs for release to GenBank by "deep sampling" of the fiber cDNA library, and 3) Normalize the arrayed cDNA library to identify 'rare' gene transcripts. The cotton fiber transcriptome, as we currently define it, consists of ~14,000 unique gene (UG) sequences that can be easily accessed as a searchable fiber dbEST (UCD Unigene/Non-redundant fiber ESTs v2.0) posted on our web site (http://cfgc.ucdavis.edu/). From the large number of UGs, it is clear that the genetic complexity of gene transcripts in developing cotton fibers is very high. Based on estimates of the number of genes encoded by the diploid cotton genome, the fiber transcriptome represents ~35% of the cotton genome! *In silico* northern analysis revealed that ~40% of the most abundant fiber transcripts representing the most highly expressed genes fall into two major functional categories – cell wall and cytoskeleton – both of which play a pivotal role in determining the size and shape of the fiber, as well as structural aspects of the cell wall that are very much a part of agronomically important properties such as fiber length. Key metabolic enzymes involved in lipid, carbohydrate and protein metabolism represent the bulk of moderately expressed genes as may be expected for developing fibers undergoing rapid cell elongation and primary cell wall biogenesis. Analysis of the fiber ESTs so far corroborates the molecular model of fibers first put forward by Wilkins and Jernstedt in 1999, and is expected to continue to provide novel insight to the complex inner workings of the cellular and biological processes required to determine a particular fiber phenotype.

Now that this phase of the EST project has been successfully completed, one of the primary aims of our initiative in cotton functional genomics at present is to focus on developing a comprehensive public database of global expression profiles. Studying the spatial and temporal expression patterns of cotton fiber genes in response to developmental, physiological, and environmental stimuli are crucial to understanding gene function, and determining how a gene's function is related to the fiber phenotype, and especially to important agronomic traits. Such information provides the foundation for increasing the sophistication and complexity of the current working model (Wilkins and Jernstedt, 1999). Microarray DNA chips fabricated from ~12,500 oligonucleotides (70-mers) has been used successfully to link genotype to phenotype in isogenic lines. For example, comparison of expression profiles of short, coarse cotton fibers relative to longer, finer fibers unveiled an altered pattern of expression in the short fibers that correlates with a decreased growth rate in which biological processes involved in fiber expansion and cell wall biogenesis are strongly suppressed. It is these types of experiments that provide us with novel insight into the genetic mechanisms underlying fiber biology, and that allows us to forge a blueprint for manipulating the fiber transcriptome towards a particular goal or direction.

The cotton fiber dbEST provides an as yet untapped resource for developing a functionally-anchored cotton genetic map. The cotton community has indicated that 5,000 DNA markers will be necessary to provide the requisite tools for genome analysis, evaluation of germplasm collections, and marker-assisted selection. At present, the number of DNA markers available in the public domain numbers in the hundreds. Although a subset of our fiber ESTs have been genetically mapped using traditional methods (http://cfgc.ucdavis.edu/), an initiative is underway to develop EST-derived SSRs as framework markers for construction of integrated consensus maps. An extensive list of potential simple and complex SSR markers has been generated from our assembled, quality-controlled EST consensus sequences for the cotton community to map EST-SSRs, which should contribute significantly to developing a high-density, functionally-anchored genetic map.

Cotton transformation and regeneration technology

Cotton biotechnology requires two processes – transformation and regeneration. At present, Agrobacterium-mediated transformation and regeneration of cotton via somatic embryogenesis remains the preferred method of choice for generating transgenic cotton, as its advantages significantly outweigh the disadvantages relative to other methods (Wilkins *et al.*, 2000). Cotton regeneration via somatic embryogenesis is highly genotype-specific, and highly regenerable lines selected from the obsolete cultivar Coker 312, serves as the industry standard at this time (Wilkins *et al.*, 2004), although linkage drag during introgression of transgenes into elite cultivars continues to be an issue of concern (Wilkins *et al.*, 2000). Major bottlenecks in cotton transformation technology are the limited number of genotypes that can be successfully regenerated, the production of high-quality callus and somatic embryos, and the low percentage of somatic embryos that successfully germinate and root into plantlets (Wilkins *et al.*, 2000).

We have recently shown that there is sufficient genotypic variability within a given cultivar to warrant screening for regeneration potential (RG) among individuals for the purpose of developing elite regenerable lines for top-producing varieties. By taking innovative approaches that break away from traditional practices, we have achieved considerable success in advancing cotton transformation technology by 1) moving closer to genotype-independent transformation by increasing the range of cotton genotypes that can be regenerated, 2) developing highly regenerable lines from elite cultivars, including California Acala cotton (Max-R), 3) decreased the regeneration time to as little as six months, 4) release of highly regenerable germplasm for immediate use and application to the biotechnology industry and forward breeding to introgress transgenes, and 5) improving transformation and regeneration efficiency to decrease production costs and accelerate commercialization of transgenic varieties (Mishra *et al.*, 2003; Wilkins *et al.*, 2004). One of the most promising technological advances we have in the pipeline is a high-throughput method of cotton transformation that is in the final stages of testing and evaluation that will revolutionize the cotton biotechnology industry in the near future.

As molecular breeding programs become increasingly reliant on marker-assisted breeding, the challenge will be to identify DNA molecular markers for RG, a multigenic trait with low genetic and high environmental variability. Three strategies are proposed for molecular breeding and genotype-independent transformation of cotton, which implemented in concert, will avoid linkage drag and dependency on one or a few related lines with agronomic properties not up to current industry standards. The first strategy recommended is a switch to our highly regenerable Max-R germplasm for development of transgenic cotton for input/output traits, now and in the foreseeable future. Strategies two and three entail introgression of RG alleles into the gene pool by using Max-R lines in breeding programs, coupled to positive selection for RG in advanced breeding lines.

Genetic modification of cotton for output traits

Genetic modification of cotton for output traits – traits that enhance food and fiber quality – is a prime target for future advances in cotton biotechnology, especially now that a new wealth of tools and resources in the form of fiber ESTs and improved transformation technologies are at hand. In our work to understand fiber gene function, de-regulated expression of some of our favorite genes in transgenic cotton has had a significant impact on yield components, such as number of seeds and fiber weight/seed, as well as fiber quality. Thus, we are establishing a database that links genes to agronomic properties, and will allow us to target manipulating expression to produce a particular fiber phenotype down the road.

One of our most promising success stories with commercial applications is a cell wall protein (expansin) that plays an important role in cell wall loosening during turgor-driven expansion of developing cotton fibers. Constitutive over expression of a fiber expansin gene in a Coker background produced a hairy leaf phenotype, and improved yield and fiber quality. This was unexpected, especially as there appears to a negative correlation between leaf hairiness and yield. In terms of fiber components measured, modest overexpression in our selected line increased or enhanced all fiber properties measured. A 4-fold increase in yield, measured as fiber weight/seed, was also accompanied by an increased number of fibers/seed. In breeding programs measuring gains in fiber length in increments of 0.01 inches, our transgenic line produced fibers 0.2 inches longer than control fibers, and increased fiber uniformity by decreasing short fiber content by ~50%. Also improved is fiber maturity, measured as a significant decrease in the number of immature fibers and a corresponding increase in the fiber maturity ratio. Not only are the gains in fiber yield and fiber quality significant in these transgenic plants, but given that many fiber properties are negatively correlated in traditional breeding programs, these welcome results underscore the vast potential of marrying genomics and fiber biology, and opens new opportunities for the genetic improvement of cotton using molecular approaches to generate novel fiber traits. We contend that only 1-2 expansin alleles in tetraploid cotton account for as much as 80-90% of phenotypic variation, and that selection of favorable alleles are unlikely to produce the degree of quantitative differences we observed in our transgenic plants. We therefore offer these preliminary results as "proof-of-concept" for biotechnological approaches to the genetic improvement of output traits in cotton.

Services for the cotton community

We invite cotton researchers to visit our web site (http://cfgc.ucdavis.edu/) for a comprehensive listing of services being provided to the community for unprecedented public access to genomic tools and resources. A brief listing of services include the following:

- Searchable cotton fiber dbEST of the UCD Unigene/Non-redundant ESTs v2.0
- UCD Unigene/Non-redundant ESTs v2.0 of ~13,000 cDNA clones
- Comprehensive list of potential EST-derived SSRs and mapped EST links
- Libraries and clones
- Printed oligonucleotide microarrays of >12,500 replicated elements, including experimental controls (projected release date 1 Sept 2003)
- EST-derived SSR database cross-indexed to mapped fiber genes

Acknowledgements

Funding from the National Science Foundation Plant Genome Program to T.A.W. is gratefully acknowledged. Sequencing of fiber ESTs was performed at facilities under the direction of Dr. Rod Wing (University of Arizona).

References

- Mishra R, H., Wang, Y., Yadav, N. and Wilkins, T.A. (2003). Development of highly regenerable elite Acala cotton (*Gossypium hirsutum* L.): A step towards genotype-independent regeneration. *Plant Cell Tissue & Organ Culture*, **73**: 21-35.
- Wilikins, T.A. and Jernstedt, J. (1999). Chapter 9. Molecular genetics of developing cotton fibers. *In* A.M. Basra (Ed.), Cotton Fibers, Hawthorne Press, New York, pp. 231-267.
- Wilkins, T.A., Mishra, R. and Trolinder, N.L. (2004). *Agrobacterium-mediated transformation and regeneration of cotton. Journal of Food, Agriculture and Environment* (in press).
- Wilkins, T.A., Rajasekaran, K. and Anderson, D.M. (2000). Cotton biotechnology. *Critical Reviews in Plant Sciences*, **19**: 511-550.