COTTON BIOTECHNOLOGY WORKSHOP: THE RUDIMENTS OF COTTON TRANSFORMATION AND BIOTECHNOLOGY

T. A. Wilkins
University of California
Davis, CA,
R. B. Turley
USDA-ARS
Stoneville, MS
N. Trolinder, L. Koonce and J. Dever
BioTex
Lubbock, TX
K. Rajasekaran
USDA-ARS
New Orleans, LA
A. Zipf
Alabama A&M University, AL
A. Nepomuceno
University of Arkansas
Fayetteville, AR

Abstract

The first commercial releases of transgenic cotton heralded the entry into a new era in cotton breeding with considerable fanfare. Although biotechnology methodology is still evolving at a fast pace, many of the questions remain the same. What gene will do what we want, and how can we control when, where and how much a transgene is expressed? How stable is the transgene and will the plant perform well in the field? These questions, and more, were the topics under discussion at the Cotton Biotechnology Workshop held during the Beltwide Conference. An overview of the steps involved in the development of transgenic cotton, including mention of experimental limitations and potential pitfalls of the technology, is presented in the text.

Introduction

The field of biotechnology is exploding at a phenomenal rate, driven by the vast potential of bioengineering for improving the overall quality of life. As techniques continue to be refined and developed, so does the opportunity to produce "designer" plants, that is, transgenic plants in which the expression of a transgene is tailor-made to suit a particular need, whether it be enhanced pest resistance, yield, fiber quality or any combination thereof. The release of the first generation of transgenic cotton for commercial production reflects the benefits, as well as the pitfalls, to be derived from biotechnology. But what is involved in the development of transgenic cotton plants? The following paragraphs address this question by presenting a general overview of key aspects of cotton biotechnology in the development of transgenic cotton as presented at the recent Beltwide meetings.

Gene Identification and Characterization

Assuming that a particular trait has been targeted for improvement, yield for instance, one of the first tasks at hand is to identify the most promising candidates from among a gene arsenal, from cotton or other species, that will produce or achieve the desired phenotype. In fact, as it turns out, gene identification and isolation is all too frequently the bottleneck to the successful development of transgenic lines with improved agronomic traits. By the early 1990's, only a handful of genes had been isolated from cotton. However, the last several years have been primarily ones of discovery, such that the number of cotton genes stored in data banks has soared to over 250 through the concerted efforts of a number of research laboratories across the country. Yet, this number constitutes but a small fraction of the thousands of genes estimated to be encoded by the cotton genome. Although genes encode proteins that perform a defined function in the growth and development of cotton, the lion's share of these genes have little, if any, practical importance from a biotechnology perspective. Thus, molecular geneticists must initially expend significant energy to isolate and characterize a gene in order to determine what function the gene serves, and assess what, if any, economic importance can be assigned to the gene.

The number of different methods or approaches for gene isolation is quite extensive - being limited only by the imagination of the investigator. A procedure called "differential screening" of a cDNA library is the most popular technique for the isolation of cotton genes. A cDNA (complementary DNA) library literally contains tens-to-hundreds of thousands of recombinant bacteria that carry copies of all the different genes that are turned on, or "expressed" in a particular tissue - fibers, for instance. Each bacterial cell contains only a single gene. Differential screening involves "plating" the cDNA library onto culture medium containing an antibiotic that allows only recombinant bacteria, that is those that carry a cotton gene, to grow. Once the bacteria have grown healthy "colonies", a duplicate set of specialized paper or "membranes" are laid on the bacterial plate and gently removed to provide an exact "copy" of each bacteria and its position on the original plate. Then membranes are chemically treated to break open ("lyse") bacteria sticking to the membrane. During this process, cDNA copies of a single cotton gene contained on recombinant DNA molecules are released, which adhere tightly to the surface of the membrane when washed to remove bacterial debris. To detect fiber-specific genes in the library, two different cDNA populations are made using a radioactive tag - one from genes expressed in cotton leaves and one from fibers. The tagged cDNAs, which serve as "probes", are allowed to bind to the recombinant DNA on the membranes in a process called hybridization. The tagged cDNA probe from fiber is hybridized to one membrane

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 1:96-99 (1998)
National Cotton Council, Memphis TN

while the leaf cDNA is used to probe the second membrane. After washing away cDNA probes that do not bind, the presence of radioactive tags that remain bound to the membrane is detected by exposing the membranes to X-ray film. Spots detected on the film indicate the position on the membrane in which a tagged cDNA probe has hybridized to a recombinant DNA molecule present in the library. The pattern of spots on the duplicate set of membranes is compared. A spot detected on the membrane probed with the fiber cDNA, but is not detected on the duplicate membrane probed with leaf cDNA, indicates a gene that is preferentially expressed in cotton fibers. This information is used to trace the newly identified "fiber gene" back to the original bacterial colony, which is isolated from the other bacteria and used to make copies of the fiber gene for further characterization. The need for such "fiber-specific" genes is based on the assumption that genes expressed only in fibers must perform a cellular function important to the growth of fibers, and hence, fiber quality.

The polymerase chain reaction, or PCR, has emerged as one of the most powerful tools for gene isolation. PCR is used to "amplify" a gene by synthesizing numerous copies of a gene, starting with as little as one molecule or copy of the gene. Differential display-PCR (DD-PCR) is a variation of the differential screening method based on the polymerase chain reaction (PCR). In this particular method, fragments of genes expressed in a particular place or time are amplified by PCR at random using a radioactive isotope or tag. The tagged gene fragments are separated by size, or "displayed" as distinct bands on exposed X-ray film. By comparing the banding patterns from two different sources (e.g. Leaves vs. Fibers), a scientist searches for the presence (+) or absence (-) of a particular band. For example, the presence of a band in fibers, but not in leaves, indicates a gene that is expressed only in fibers, and therefore, may be important for fiber improvement.

Many of the methods currently in use for gene identification, in fact, rely on this type of +/- screening, but are not necessarily restricted to the use of nucleic acid (DNA or RNA) probes to detect differences in gene expression. A more traditional approach is to compare differences in proteins, the gene products that are directly responsible for determining the size and shape of a fiber, and hence, fiber yield and fiber quality. Once identified, the protein is partially purified and can be used in a variety of ways to develop probes that can, in turn, be used to identify the corresponding cDNA or gene that encodes the protein. Although it takes longer to isolate the genes, a protein-based method is oftentimes the only means available. However, low-abundant or rare proteins generally fall below the detection limits of this method, and a nucleic acid approach is generally the preferred method of choice under these circumstances.

The bulk of cotton genes isolated to date have been discovered through the type of global searches described above, often referred to as "fishing expeditions" since the screen depends on the search for +/- differences at random. As we begin to expand our knowledge of cotton physiology at the molecular and cellular levels, we can begin to focus our search for key genes, and target a particular gene for isolation. For instance, yield can be theoretically increased by any number of possible routes, including the genetic manipulation of genes that impact carbon sources, carbon sinks, nutrient uptake and utilization, water use efficiency, increased stress tolerance and pest resistance, etc. In the short-term, changing the expression of one or more of these genes may increase yield potential or yield. For long-term goals, it is imperative that researchers be able to control and manipulate the expression of key genes in each of these cellular processes as a group to produce the most beneficial combination of genes. Mapping genes of known function and determining the best possible combination of genes will also be important to breeding programs using marker-based selection.

Once the gene of interest has been identified, isolated and purified, characterization of the gene begins. DNA sequencing is the process by which the gene sequence, or the unique order of nucleotides (DNA building blocks) that make up a gene, is determined. The sequence can be used to search for similarity to related genes reported from other species that are maintained in data banks. High similarity scores to a gene of known function in another species helps in assigning a possible function to the cotton gene. This knowledge assists in determining how the gene can used to its full potential to mediate a positive change in the phenotype of transgenic plants, in terms of desirable agronomic traits, plant stature and architecture. Gene characterization analysis also entails conducting a series of gene expression studies designed to determine when and where the gene is expressed, and how the expression pattern is related to the function of the gene. At this point, a strategy is planned for changing the phenotype of the cotton plant (e.g., yield) by manipulating the expression of the gene in transgenic plants.

Expression of Transgenes in Transformed Cotton Plants

The next step in the evolution of a transgenic plant is the assembly of a transgene. The transgene is actually an artificial gene assembled from parts of different genes and is therefore referred to as a chimeric gene or expression cassette. The different components of a gene that are an absolute necessity for the transgene to function are: the promoter, the coding region of a gene, and a termination sequence. The gene promoter functions as a cellular "switch", in that it regulates when, where and how much the gene is expressed. The coding region is the DNA sequence that determines what kind of protein is made in the plant cell. The termination sequence serves as a "stop" signal to tell the cell that this is the end of the transgene. These three components are "glued" together and maintained as a

recombinant DNA molecule in bacteria (*E. coli)*. The promoter most often used is the strong, constitutive 35S promoter from the Cauliflower Mosaic Virus (CaMV), which is transcriptionally active or expressed at all times in all cells and tissues.

For long-term studies, an extensive collection of tissue-specific promoters will be needed to limit expression of a transgene to a particular tissue or cell. At present, only a meager number of cotton promoters are available, primarily because of the amount of work required to analyze a promoter before it can be used for the assembly of transgenes. One means of testing promoter activity is by fusing the promoter to a reporter gene, such as GUS, luciferase or the jellyfish green fluorescence protein. Any gene that is not normally found in plants and is easily detected can be used as a reporter gene. The expression of these reporter genes can be visualized in transformed cells and plants because of their ability to fluoresce or develop characteristic color signatures in the presence of an exogenous substrate, and revealing when and where a gene is expressed as a result. As an example, the GUS gene is fused to a promoter and introduced into cotton. When the reporter gene is expressed, it produces the enzyme β-galactosidase which converts a colorless chemical to a blue colored product. Thus, if only fibers turn blue when treated with the colorless chemical, then the promoter "switches-on" expression of the reporter gene only in fibers, and is said to be a "fiber-specific" gene promoter.

Particle gun bombardment is oftentimes used to ensure that a promoter is functional, and therefore capable of driving the expression of a transgene. This approach involves coating small tungsten or gold particles with the DNA of the promoter-reporter gene construct and literally shooting a target tissue (leaf, embryo, callus, etc.) with these particles under high velocity and in a vacuum. Detection of the reporter gene in bombarded cells indicates a functional promoter and the chimeric reporter gene construct can be introduced into cotton to determine where the gene is normally expressed in the plant and how strong the promoter is in terms of how much reporter gene is present. Intact transgenic cotton plants have also been successfully obtained using particle bombardment procedures, although transformation is significantly less efficient that *Agrobacterium* methodology.

Construction of Binary Vectors

The transgene expression cassette is isolated from recombinant DNA molecules maintained in bacteria and cloned into a DNA molecule called a binary vector, which provides the information necessary for introducing a transgene into a plant. The binary vector carries DNA normally found in natural populations of *Agrobacterium tumefaciens*, the soil-borne bacteria that causes crown-gall disease in plants. The regions flanking the T-DNA (transferred DNA) on the right and left borders carry the DNA sequences responsible for the transfer of the T-DNA. The region between the border sequences of the T-DNA has been replaced with another chimeric gene that encodes an antibiotic resistance gene (e.g., kanamycin) that is used as a selectable marker for identifying cells transformed with the transgene. The expression cassette is cloned into the region between the T-DNA border sequences as well. The T-DNA, carrying the selectable marker and the transgene, eventually becomes incorporated into the DNA of the plant, and thereby becomes part of the plant's genetic make-up or blueprint (genotype).

Any alteration or deviation in expression of a gene beyond what is normally expected is called ectopic expression, and is therefore a term that often applies to transgenes. The expression of the transgene can be used to modify the level and/or timing of expression of endogenous genes, those genes normally expressed in the plant, to induce the desired change in phenotype. Several strategies are currently available to accomplish this goal, although how each one works at the molecular level is not necessarily understood. One approach is simply to "overexpress" the target gene by introducing a transgene under the control of the same or a different promoter. Another approach, called "gene silencing", uses the transgene to effectively "turn off" the expression of an endogenous gene. Co-suppression is a type of gene silencing, in which the expression of the "wild-type" or normal endogenous gene is hampered in some way by the high level expression of the transgene. Thus, overexpressing a transgene may lead to unexpected results. Another gene silencing phenomena relies on cloning of the transgene in the "antisense" orientation in the expression cassette. In this case, the coding region of the gene is cloned in backwards to generate an "antisense" gene transcript when the transgene is expressed. The transgene itself codes for nonsense, but when expressed, the antisense gene transcript interferes with the expression of the wild-type gene in transformed plants, and reduces or completely eliminates the amount of protein produced by the wild-type gene, leading to an altered phenotype in the absence of an important protein.

The current state-of-the-art is geared for manipulation of single gene traits, and we cannot as yet effectively manipulate complex genetic traits under the control of multiple genes. Since most agronomic traits are, in fact, complex traits, one important task for the future is the development and implementation of multigenic expression cassettes that will allow the introduction of more than one gene at a time. Until then, conventional genetic crosses are being performed to introduce transgenes for different traits (e.g., herbicide and insect resistance) into a single plant in a scheme called gene pyramiding. More recent introductions, however, carry both herbicide and insect resistant genes in a single expression cassette, in which the herbicide resistant gene is used as the selectable marker for identifying transformed plants in place of antibiotics.

Agrobacterium-Mediated Transformation of Cotton Hypocotyls

The successful development of transgenic plants is absolutely dependent on the ability to regenerate an entire plant from undifferentiated cells called callus. The unique ability of plants to regenerate is due to the retention of all genetic information in each cell, a property called totipotency. However, while this is true in theory, not all plant species regenerate with similar ease and are referred to as recalcitrant plants as a consequence. Cotton is one of the more notorious of the recalcitrant plant species. Despite this characteristic, researchers have been successful in regenerating cotton, although this success is limited to only a very few upland cultivars and is therefore genotype-dependent at this time. An elite regenerable line of the upland cultivar Coker 312 currently serves as the industry standard for *Agrobacterium*-mediated transformation of cotton. The transfer of transgenes into different genetic backgrounds is readily accomplished via selection for an active transgene in a conventional backcross program.

From start to finish, it takes approximately 8 to 10 months to recover a transgenic plant. The whole process begins by germinating seedlings for about 10 days under sterile conditions in the laboratory. The stem-like region below the leafy cotyledons, called the hypocotyl, is cut into small pieces that become the 'explants' in tissue culture terms. After each explant is coated in a solution of *Agrobacterium* that carry the transgene in the binary vector, it is incubated in the dark for 48 to 72 hours for "co-cultivation". It is at this point that *Agrobacterium* infect the explants at wounded sites, the T-DNA is transferred from the bacteria into the plant cell and becomes incorporated into the plant DNA.

The next phase of the transformation process involves the formation of callus. Following co-cultivation with *Agrobacterium*, the explants are transferred to tissue culture medium containing the antibiotic kanamycin or other selectable markers (e.g., herbicide), depending on the resistance gene used. Under these conditions, only cells that have been 'transformed', that is they carry the transgene and the selectable marker, multiply and grow to form callus. The *Agrobacterium* and untransformed cells die in the presence of kanamycin.

Once healthy callus is formed, a combination of plant hormones is added to the culture media to induce the formation of embryogenic callus. Once formed, embryogenic callus is put into liquid suspension culture without hormones to promote further development and the rapid formation of somatic embryos in large quantities. The somatic embryos are equivalent to seed embryos, but without the seed coat at this stage. The immature somatic embryos are subsequently transferred back onto solid medium to allow further development of the embryos to take place. In order to mimic seed development as closely as possible, mature embryos are dehydrated on media that removes moisture from the embryos in a way that simulates normal seed dessication. The dehydrated embryos are finally ready for germination; the embryos are transferred to rooting medium and placed under high light to facilitate the formation of roots and shoots. After small plantlets have been allowed to grow sufficiently, cuttings are made to produce a number of identical "clones". Several of the rooted clones are potted in soil and gradually "hardened off", that is they are gradually exposed to the outside environment to allow the plants to become acclimatized to the lower humidity. These plants, called the primary transformants or the T_0 generation, are grown to maturity in a greenhouse environment to produce T_1 seeds.

Analysis and Characterization of Transgenic Lines

It is at this stage that the real analysis begins to confirm that the kanamycin-resistant plants regenerated carry the transgene, but more importantly, actually express the transgene and produce the desired phenotype. Because not all transgenic plants express the transgene to the same degree, it is therefore essential that a collection of independent transgenic lines be produced for each gene construct to be introduced, and that each transformed plant be thoroughly evaluated before being released. Another key factor that must be kept in mind is that plant chromosomes exist in pairs and genes are normally present in two copies, one gene on each of the homologous chromosomes. In contrast, however, T_0 plants are genetically "hemizygous", that is the plants carry only one copy of the transgene on only one of the homologous pair of chromosomes. This means that the T_0 are genetically "unstable" and T_1 progeny produced from the T_0 plants will segregate to produce progeny bearing either two copies, one copy or 0 copies of the transgene. It is there essential that the transgene is genetically stabilized so that is transmitted to all offspring in subsequent generations, and moreover, that selection pressure is maintained to ensure that transgene continues to produce the desired phenotype.

Both genetic and molecular analysis of transgenic lines are further complicated by the fact that, on average, multiple copies (1-6) of the transgene are inserted at random in the plant genome. Yet, each transgene may contribute to the phenotype to varying degrees, or not at all, depending on where the gene is inserted in the genome - a phenomenon called positional effect. This situation stresses the importance of screening a sufficient number of plants in segregating progeny populations to select plants with the fewest number of transgenes that show the desired phenotype for the following reason. The transfer of transgenes from Coker 312 to elite cultivars is currently managed through a backcross program, and increasing the number of transgenes increases the difficulty in selecting plants for cultivar development. To alleviate some of these problems and to ease the way through the regulatory process in the commercial release of bioengineered plants, most

companies demand that transgenic plants contain only a single transgene, without any rearrangements, and no DNA flanking the T-DNA be present. Genetic recombination is yet another factor that can profoundly impact the expression of the transgene from generation to generation. It is an unfortunate, but documented fact, that transgenes can be "silenced" over time. It is therefore imperative that the plants be monitored in each generation to ensure that the desired phenotype is maintained.

Summary

The first commercial releases of transgenic plants in the last year show that cotton biotechnology has come a long way in a relatively short period of time. Clearly, the increasing number of transgenic lines in variety tests across the country indicates that biotechnology will be a key factor in cotton breeding programs well into the next millennium. The field of biotechnology is still in its infancy, and as techniques continue to be developed and refined, so too will our capacity to produce "designer" cotton plants for improved agronomic performance and high quality fiber. This should be highly welcome, especially in light of increasing concerns about yield plateaus, an increasingly narrow germplasm base, rising production costs, and population growth.