DNA MARKERS: A NEW TOOL FOR IMPROVING THE BREEDING OF AUSTRALIAN COTTON CULTIVARS

Dainis Rungis[1], Danny Llewellyn[1], Elizabeth Dennis[1] and Bruce R. Lyon[2]

CRC for Sustainable Cotton Production, [1]CSIRO Plant Industry, Canberra, [2]University of Sydney, Sydney

Introduction

When breeders cross two cotton varieties together there are millions of potential new combinations of genes that are generated in the progeny from the two parent plants. The breeders must then select from amongst these those few combinations that give an improvement in varietal performance (eg. yield, fibre quality, disease tolerance). In many cases this requires the assessment of many thousands of lines in small evaluation plots over multiple seasons to come up with those elite performers that are then evaluated in large scale variety trials and a selection of which will become a new variety. Any new tool that will help the breeder follow useful agronomic characteristics through a breeding population and reduce the number of lines that need to be evaluated in the field will increase the efficiency and speed with which new varieties can be developed. DNA or molecular markers are one such tool that is now being used widely for fundamental genetic studies in plant biology, but increasingly in commercial and institutional breeding programs for crops, such as maize and wheat (where incremental advances in yield have become difficult to achieve by conventional strategies). DNA markers have yet to have any major impact on cotton breeding.

DNA markers are simply differences in the DNA sequence between different individuals that can be used as genetic tags for genes since, like genes themselves, they are passed on to offspring in a normal Mendelian inheritance. They can be as simple as single substitutions of one base in the DNA for another base, deletions of parts of the DNA or additions of DNA sequences in that particular region of the chromosome. These DNA sequence differences can either be within genes or next to genes and are detected by simple physical assays for the size of a band on a gel, for example. Following the inheritance of the DNA band (marker) in different progeny plants is functionally the same as following the inheritance of the DNA sequence difference and hence of the gene(s) within or adjacent to which they occur. The DNA marker does not have to be responsible for a genetic

character that it is used to follow, but must be tightly linked to the gene(s) if it is to be of any use in breeding. As an example of the utility of DNA markers, let us say that there are a number of major genes in cotton that together are responsible for the high fibre quality of an Australian variety and that we have been able to find three or four DNA markers that most of the time occur in any cotton variety with high fibre quality, and similarly a few markers have been found that correlate with tolerance in another variety to Verticillium wilt. If a breeder were trying to combine the high quality of the one variety with the disease tolerance of the other then the crosses could be carried out and instead of taking a large number of lines through to F4 or F5 generations to bulk up seed and testing them all for both quality and disease tolerance in reasonably large experimental plots, they could simply test individual F2 plants for the presence of the DNA markers for both quality and disease tolerance and only keep those that possess all markers. Such lines would have a high probability of containing both quality and disease tolerance attributes. This assessment could be done as soon as the seedlings appear out of the ground by taking a small piece of the cotyledon and scoring the presence of the markers in some DNA isolated from each of a few hundred individual plants. The breeder might then get away with only bulking up a dozen or so plants for large scale testing instead of the hundreds or even thousands that might need to be tested using current breeding strategies. DNA markers therefore provide a powerful tool for crop improvement because of this ability to quickly track the presence of both single genes and quantitative characters (like yield) in individual plants at an early stage, even before the character is expressed.

DNA markers are simple physical tags for potentially complex genetic traits and can be used in both breeding programs and in gene-mapping studies just like traditional genetic traits like okra leaf or petal colour, for example, but, because they represent just small random differences in DNA sequence that may have no observable effect on the shape or properties of the plant many thousands of markers will exist in any one plant. It should, therefore, not be too difficult to find a DNA marker near any gene(s) of interest. They will have particular utility in the study and breeding of disease tolerant plants where the classification of the tolerance of individual plants is often difficult and inconclusive and where we are only just learning how these genes function to confer disease tolerance. In addition many agronomically important traits such as yield and water use efficiency, for example, are controlled by more than one gene and are difficult to estimate from single

plants, which makes selection of improved breeding stocks through these traits difficult. By using molecular markers to map where genes are on chromosomes, the presence of those desirable genes, or at least the genes that contribute a major part of complex traits like yield, can be detected earlier and with greater certainty than by using the traits directly as selection criteria. The availability of a large number of DNA markers spread throughout a plant's genome could also have important uses in speeding up the introduction in a back-cross breeding program of transgenic traits, like insect tolerance, from a poorly adapted cultivar into elite cultivars by allowing the selection of breeding lines containing mostly the elite genome at an early stage of the breeding process. Molecular markers have other uses such as "fingerprinting" cultivars for Plant Breeder's Rights registration and to assess the genetic relatedness between cultivars. This latter characteristic is often useful in crop species where hybrids are used, in order to determine the most genetically dissimilar cultivars in an attempt to predict the most successful hybrids. Markers can also be used to construct detailed maps of regions surrounding particular genes, and then those maps can be used to clone the genes and advance our understanding of how genes function in plants to confer particular phenotypes or traits.

Molecular marker techniques are a new development in plant science, but as indicated above, they provide a very powerful tool for gene analysis and breeding that is already being exploited with great success in both cereals, such as rice, maize, wheat and barley and in dicot crops, such as soybean, canola, lettuce and tomato. It is, however, often difficult to apply to cotton many of the molecular and biochemical analyses that are routine in other plants, because of cotton's high chemical content (eg. gossypol), and many of the existing marker technologies will only be transferred to cotton with some difficulty. A number of groups are actively investigating the use of DNA markers for cotton improvement and some progress is being made, at least at a scientific level. If the Australian cotton industry is to remain competitive internationally, and is to take advantage of the possibilities offered by molecular genetics, then the fundamental techniques of molecular genetics and marker isolation and use must be transferred to cotton within the context of the CSIRO cotton breeding program that currently generates the bulk of the new cotton varieties for the Industry. We have begun a new research program to look at the potential for isolating and using specific kinds of DNA markers in cotton and hope to extend those techniques to our breeding team if useful markers become available from this

or other research projects.

Types and uses of DNA markers

The genetic markers first used for breeding and the study of inheritance were morphological markers. These are physical traits often visible to the naked eye. Genes themselves determine these traits, and so in effect, a plant character is being used to tag the gene. Examples of morphological markers are traits such as leaf shape, flower colour, dwarfism etc. While these traits are not very abundant they are useful for genetic studies, but are rarely now useful for breeding programs as the morphological marker phenotype is often deleterious to the plant, and is not desirable in elite breeding material. The second type of markers used were isozyme markers that rely on detecting changes to genes for enzymes that result in the production of a modified enzyme (or isozyme) having a different physical property, such as how quickly it migrates through a gel in response to an electric current. The DNA sequence changes in the gene result in the substitution of one amino acid for another in an enzyme and if this involves a difference in electric charge the migration of the enzyme can be either accelerated or retarded in the gel and this can be visualised by specific chemical stains that can detect the presence of the enzyme. The variant isozyme markers can then be followed in breeding populations by making small extracts from the plants and running them on gels and staining for the different patterns of isozymes characteristic of the two parents. Although often more numerous than morphological markers, there are not large numbers of different isozyme markers in plants, because many changes in the sequence of a gene can deleteriously affect how well an enzyme works and these would reduce the viability of the plant.

DNA markers, on the other hand, have many advantages over morphological and isozyme markers. The primary one being that, because they are based on random differences or polymorphisms in the DNA sequence between two cultivars or species, there are huge numbers of them in any plant, so markers can always be found which are very close to, or tightly linked to any particular gene(s). While these differences in the DNA may alter the functioning of a gene, and so give rise to different traits or morphological phenotypes, most of the time these differences at the DNA level are neutral and will not be detectable unless the DNA is physically examined for the polymorphisms. Molecular markers can be

detected as soon as DNA can be extracted from a plant and in most cases this can be very soon after germination, or even before, if part of a seed can be removed without affecting germination. This means that a plant does not have to mature before the presence or absence of a trait can be assessed. This makes it possible to track gene(s) responsible for a trait with greater certainty, and also makes it easier for the gene(s) to be cloned using map-based cloning techniques.

There are several different molecular marker techniques and new protocols and variants are continually being developed. These include methods such as, RFLP's, RAPD's, AFLP's and SSR's. Each is a unique method sampling different types of sequence variation, with its own advantages and disadvantages and level of sensitivity for detecting DNA sequence differences between individuals or cultivars. The oldest, RFLP's (Restriction Fragment Length Polymorphisms) involves cutting genomic DNA at defined sites using bacterial enzymes that recognise specific DNA sequences. Any changes in the DNA sequence at the cut site prevent the cleavage of the DNA and this will alter the length of the cut sequences to the next adjacent cut site. The size of the fragments are detected using DNA hybridisation to a specific cloned gene sequence so basically you are detecting sequence changes at cut sites immediately around a specific gene(s). This method has been used extensively in many studies and species, but has the drawback of requiring large amounts of DNA per reaction and the use of expensive and complex radioactive or non-radioactive detection systems making it impracticable for large-scale applications in plant breeding.

The remaining methods rely on the PCR (polymerase chain reaction) technique. This is a process by which very small amounts of DNA can be exponentially amplified to a point where it can be clearly seen on a simple agarose gel. This means that only a small amount of starting DNA is required for each reaction, and so it is practical using these methods to process many hundreds of individual DNA samples isolated from say a small clipping of a leaf or cotyledon. RAPD's (random amplified polymorphic DNA) utilises the PCR reaction to amplify random regions of the genome defined by short random sequences or DNA primers (Welsh and McClelland, 1990 ; Williams et al, 1990). When these sequences bind to a region of the DNA and are not too far apart they will amplify a DNA band of a particular size. Sequence differences between two individuals at the site at which the primers bind or small insertions or deletions between the sites may show up as bigger or

smaller bands, or no band at all, that can then be used as the physical marker in segregating progeny to tag or follow a gene. This method is highly dependent on reaction conditions, and so can be difficult to repeat, especially in different laboratories and has proven to be difficult to transfer between researchers. Much more of the genome is sampled for DNA sequence changes with each primer combination than with RFLP's, but many hundreds of primers need to be tested on the two parents in a cross to identify polymorphisms and the markers are often not transferable between different crosses of the same species. AFLP's (amplified fragment length polymorphism) is another PCR based marker technique that also involves cutting genomic DNA at defined sites, as with RFLP's, but the fragments produced are amplified using the PCR technique (Vos et al, 1995). This method has been developed fairly recently, but already it has been used in countless mapping projects in a wide variety of plant species. The main advantage of this method is that it generates a large number of marker loci per individual reaction (50-100) compared with 1-5 for RAPD's. This means that many potentially polymorphic markers can be surveyed in a relatively short period of time making it useful for fine mapping of chromosome regions or for finding polymorphisms in species or populations where the level of genetic polymorphism is quite low. Another advantage of the AFLP technique is that unlike RFLP's, it does not require any prior sequence knowledge of a genome before it can be applied. The final technique is SSR's (simple sequence repeats) or microsattelite markers. Microsatellites are short simple sequence repeats (like GAGAGAGAGA) scattered throughout the genome, which are often highly variable in length due to the loss or gain of a number of repeating units. They were first developed for use in mammalian systems, but have now been found in all animals and higher plants. The length of the repeating unit is detected by amplifying the region across the repeat with specific unique primers flanking the repeat. Because they are highly variable, these repeats are ideal as markers and can often be transferred between different genetic crosses and are slowly replacing traditional morphological markers in many genetic systems. The disadvantage of these SSR markers is that they must be developed specifically for the biological system in which they are to be used. However, the high utility of microsatellite markers in many different breeding and genetic studies means that the initial effort to develop them in a particular system is rewarded in the longer term.

The initial object of any gene mapping project is therefore to use markers of one sort or another to find the small DNA sequence differences that exist around target genes in the

two lines being crossed. Following these markers through the progeny from different crosses and observing their inheritance, the recombination that occurs between them, and their co-inheritance with particular agronomic traits allows the markers to be placed in a linear order to create a genetic or physical map of a genomic region that can include both DNA markers, any known isozyme or morphological markers and agronomic traits. These maps can span the entire genome, where many hundreds of markers may be required to give uniform coverage, or they can be restricted to small regions where just a few markers can create a detailed map that will then allow the tagging of an important agronomic trait or the isolation and characterisation of a particular gene(s). Map-based cloning of genes relies on finely mapping the region around a particular gene, and then using this map to locate the gene on a small fragment of the genome that can easily be sequenced, while marker-assisted breeding involves finding markers tightly linked to a particular gene of agronomic importance, and then using these markers to track the presence of the gene through the breeding process (Staub et al, 1996).

DNA markers in cotton

Molecular marker techniques have not been widely used in cotton, in spite of it being such an economically important crop species in many parts of the world. However, cotton has benefited from conventional breeding programs, which have resulted in yield increases of 7-10kg lint/ha/yr since 1934 (Meredith, 1994). If molecular markers could be successfully developed in cotton, then selection in breeding programs could be hastened, and routine molecular genetic techniques for the analysis and cloning of genes could be more easily applied to cotton. The now widespread use of transgenic traits in cotton has also prompted a re-evaluation of marker-assisted breeding to accelerate the breeding of elite transgenic cultivars.

The genus *Gossypium* has about 50 members, many of which are native cottons and wild diploid species. Of the commercially cultivated species, *G. hirsutum* or upland cotton comprises about 90% of the world crop, and *G. barbadense* or Pima cotton fills the remaining demand for higher quality fibres. So far, molecular markers have only been used to "fingerprint" cultivars within *G. hirsutum*, and genetic map construction has been restricted to some interspecific crosses between *G. hirsutum* and *G. barbadense*. The inattention to the use of DNA markers in upland cotton improvement appears to be largely

because there are not large numbers of DNA markers that can be detected within the *G. hirsutum* cultivars. Many more DNA sequence differences will obviously exist between two different species than within a species, but in many other crops there are still quite high levels of DNA sequence variation within and between cultivars that allows many DNA markers to be found and utilised. A genetic map constructed using an interspecific cross relies on the higher sequence variation between species and ensures that many markers can be found that are linked to any particular traits. In cotton improvement there are some useful traits, mainly involving fibre quality, that are being transferred from *G. barbadense* into *G. hirsutum* that could be enhanced by marker-assisted breeding and this appears to be one of the main current uses of the technology. However, if this type of marker breeding is to be used in upland cotton, a method of detecting markers and mapping genes from crosses between different *G. hirsutum* cultivars must be developed, as they form the majority of commercial cotton cultivars grown in the world today and consequently the majority of the crosses being carried out by breeders. When mapping a particular gene(s), only markers closely linked to the gene(s) of interest are useful so if the level of genetic variation within *G. hirsutum* cultivars is very low, as we expect, then polymorphic markers linked to a particular gene in upland cotton might be extremely rare.

All of the modern molecular marker methods have recently been applied to *Gossypium* species with varying degrees of success, usually dependent on the mapping crosses being used. As indicated previously, most marker work in cotton has been done with interspecific crosses – usually between *G. hirsutum* and *G. barbadense* or between different diploid species. Little has been done on mapping markers within *G. hirsutum* crosses other than to distinguish different upland cultivars or to estimate genetic relatedness, but generally the level of polymorphism is very low compared with that seen in other crop species. The RAPD technique has been applied to several Australian cultivars, and up to 8% of bands were found to be polymorphic between cultivars (Multani and Lyon, 1995). With AFLP's, one of the techniques most sensitive to DNA sequence variation, studies have found between less than 1% to 6% polymorphism between cultivars (Akbari et al 1998 ; S. Saha, pers. com.; D. Rungis, unpublished results). While these rates of polymorphism are sufficient to easily distinguish cultivars, the chances of finding a polymorphism close to a particular target gene are small and a lot of marker loci will need to be assayed to find useful markers which are tightly linked to a particular gene(s). A set of about 200

microsatellite markers has been developed in cotton from a *G. hirsutum* and *G. barbadense* cross (Benjamin Burr, unpublished results), and while they can detect abundant polymorphisms between *Gossypium* species, they exhibit a very low level of polymorphism between *G. hirsutum* cultivars (D. Rungis, unpublished results) and are unlikely to be useful for upland cotton breeding except for the introgression of genes from Pima cotton.

The project being developed at CSIRO as part of a post-graduate studentship with the CRC for Sustainable Cotton Production and the University of Sydney involves the development of molecular markers in cotton, and particularly in ascertaining if marker-assisted breeding and map-based cloning is feasible in upland cotton. As part of this project, we are attempting to map the location of two genes that have an easily scored visible phenotype - the gene for okra leaf, and the gene for brown lint. These genes will serve as model systems to allow us to develop all the necessary protocols for isolating and using DNA markers in cotton and will also provide us with some interesting science on how cotton plants control the development of leaf shape and fibre colour. The mapping population used for this study is the F2 generation of the cross between Siokra V-15 (okra leaf, white lint) and PD93004 (normal leaf, brown lint) an American upland cotton. The two traits for leaf shape and fibre colour segregate independently in the F2 of this cross and this allows the one cross to be used to map and clone both genes. Preliminary studies with both AFLP's and microsatellite markers indicate that the level of genetic diversity between the different upland cotton cultivars, including an Australian cultivar is very small, even though they are of very different pedigree and this may make it very difficult to find closely linked markers to the two target genes. Direct sequencing of random parts of the genome of three different upland cultivars of diverse pedigree has confirmed that very few DNA sequence changes are present between upland cultivars. These data all suggest that cultivated upland cotton has in the not too distant past been through a very narrow genetic "bottleneck" that makes them very uniform genetically and this is consistent with the historical accounts of the development of our modern cotton cultivars. Further studies will be carried out on screening a larger number of AFLP primer combinations to find those rare markers that might be linked to either okra leaf or brown lint genes and to fully evaluate the potential for using DNA markers in cotton breeding, but initial indications are that good DNA markers will be difficult to find in cotton and will have to be for high value

traits like tolerance to Verticillium or Fusarium wilts to warrant the resources and effort that will need to go into finding them in the first place.

Acknowledgments

This project is supported by the CRC for Sustainable Cotton Development. The gift by Peter Reid of F2 seeds used for this study is gratefully acknowledged.

References

Akbari M, Kota R, Becerra Lopez-Lavalle LA, Lyon BR (1998) Verticillium and Fusarium wilts of cotton: Molecular genetic markers for disease resistance. Proceedings of the Ninth Australian Cotton Conference, Broadbeach, Australia (submitted)

Meredith WR (1994) Use of molecular markers in cotton breeding. pp.303-308 *In* Constable GA and Forrester NW (eds) Challenging the future: Proceedings of the World Cotton Research Conference I, CSIRO, Australia.

Multani DS and Lyon BR (1995) Genetic fingerprinting of Australian cotton cultivars with RAPD markers. Genome 38:1005-1008

Staub JE, Serquen FC, Gupta M (1996) Genetic markers, map construction, and their application in plant breeding. HortScience, 31:729-740.

Vos P, Hodgers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407-4414

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213-7218

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531-6535