Final Report

Part 1 - Summary Project Details

Report Due Date: 29-Sept-00

CRDC Project Number CSP85C

Project Title:
(< 15 words)
Development of genome specific markers for the Australian C, G, and K genomes and continuing development of intergenomic germplasm

Part 2 - Project Contact Details

Admin Contact: Ms (Title (ie Mr/Ms)) Dianne (First Name) Rosson (Last Name)

Organisation: CSIRO Plant Industry
(Name of organisation that will be administering the funding)

Postal Address: GPO Box 1600 (PO Box) (Street)
Canberra (Town) ACT (State) 2601 (Postcode)
02-6246-5277 (Phone) 02-6246-5000 (Fax) d.rosson@pi.csiro.au (Email)

Principal Researcher: Dr. (Title (ie Mr/Ms)) Curt (First Name) Brubaker (Last Name)

Organisation: CSIRO Plant Industry
(Name of organisation that will be administering the funding)

Postal Address: GPO Box 1600 (PO Box) (Street)
Canberra (Town) ACT (State) 2601 (Postcode)
02-6246-5085 (Phone) 02-6246-5000 (Fax) curtb@pi.csiro.au (Email)

Supervisor: Dr. (Title (ie Mr/Ms)) Tony (First Name) Brown (Last Name)

Organisation: CSIRO Plant Industry
(Name of organisation that will be administering the funding)

Postal Address: GPO Box 1600 (PO Box) (Street)
Canberra (Town) ACT (State) 2601 (Postcode)
02-6246-5081 (Phone) 02-6246-5000 (Fax) tonybr@pi.csiro.au (Email)

Part 3 - Final Report

1. Outline the background to the project.

The indigenous Australian *Gossypium* species represent an underutilized and potentially valuable resource for cotton breeding in Australia. Some of the species appear to be resistant to *Fusarium* pathotypes found in Australia, and some have gossypol-free seeds. Both of these traits would useful to cotton breeders if they can transferred to commercial cotton cultivars. Transferring genes from the wild species into cultivated cotton, however, is difficult. The indigenous *Gossypium* species have fewer chromosomes than the cultivated cottons and hybrids between the two are sterile. Furthermore, because the chromosomes of indigenous species and those of the cultivated cottons differ genetically and structurally, exchange of genetic material is limited. Consequently, transferring genes from the wild species to the cultivated cottons requires means to overcome both the hybrid sterility and the low levels of chromosome recombination.

This project continues and extends the development of intergenomic germplasm initiated under CSP47C. Under CSP47C, fertile synthetic hybrids that incorporate chromosomes from indigenous *Gossypium* species into cultivated cotton were developed, overcoming the fertility barrier. These new fertile hybrids include indigenous species that have gossypol-free seeds and possible resistance to *Fusarium* wilt. Under that grant a variety of strategies for transferring genes from the native species were tested. In the end, a process called hexaploid bridging was identified as being the most likely to maximize the number of exchanges between the chromosomes of the indigenous *Gossypium* species and those of the cultivated cottons. This process involves repeatedly backcrossing the synthetic hybrids to cultivated cottons. Over several generations the chromosomes of the native species are progressively lost, but in the process some of their genes are left behind on the cultivated cotton chromosomes.

Having overcome the fertility barrier, the challenge is to overcome the low frequency of genetic exchange. Traditionally, transferring genes between two interfertile species is achieved by planting out large numbers of hybrid plants and selecting those that express the desired characteristic. This is a highly effective strategy, but unfortunately a process that is ineffective in this situation. Firstly, although the hybrids developed under CSP47C are fertile, they do not produce sufficient seed for large-scale field trials. Secondly, plants that express the desired characteristic will do so for one of two reasons. They either still have the wild chromosome carrying the gene of interest or they are recombinant individuals and carry the target gene on a cultivated cotton chromosome. It is the latter group of plants that are desired. Once these recombinant individuals are identified, they can be transferred to standard breeding programs. The purpose of CSP85C was to develop molecular markers that allow us to identify the chromosomes carrying the genes controlling the genes of interest and to differentiate between recombinant individuals and those still carrying the desired genes on the chromosome of the indigenous donor species. For the purpose of the grant the inheritance of *Gossypium australe* chromosomes was the focus. This species has gossypol-free seeds.

2. List the project objectives and the extent to which these have been achieved.

The specific objectives as detailed in the grant proposal and the extent to which they have been realized is detailed below.

Develop molecular markers.

After careful consideration AFLPs were selected as the marker system of choice. The advantages of AFLPs are their reliability, efficiency, and the fact that only small amounts of starting DNA are required. While developing the mapping population, a small pilot project in which 35 *G. sturtianum* accessions were screened for genetic polymorphisms at 464 loci (using 14 primer combinations) (Fig. 1). This permitted a thorough evaluation of AFLPs as a marker system and allowed us to optimize the AFLP protocols.

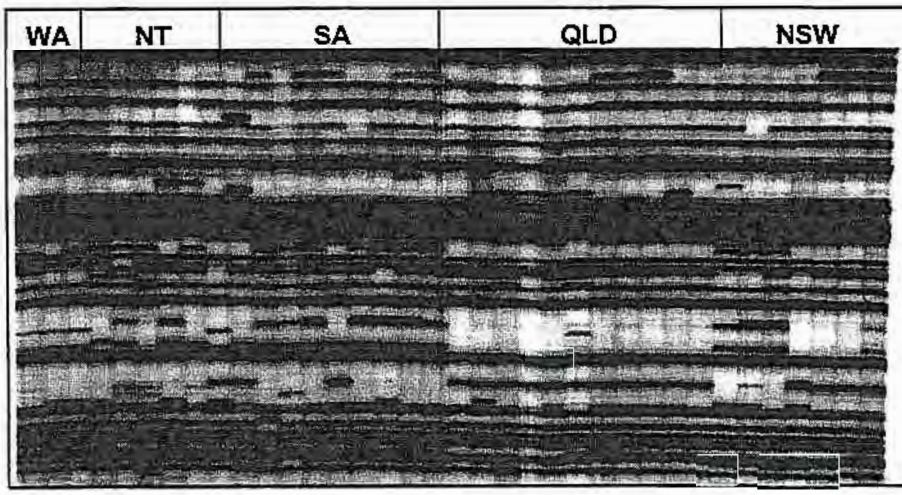


Figure One. Partial AFLP profile of 35 *G. sturtianum* accessions using primers *Eco*R1-AAG and *Mse*I-CAT.

(WA=Western Australia; NT=Northern Territory; SA=South Australia; QLD=Queensland; NSW=New South Wales)

Synthesize mapping populations

A primary goal of this project was to construct the first genetic linkage map of an Australian *Gossypium* cotton genome. To do so, two segregating populations were used. The first is a standard genetic mapping population of F2 progeny derived from an interspecific F1 hybrid. The second is a population of 18 BC1 aneuploid individuals derived from a cross between a pentaploid individual that carries a single copy of the *G. australe* G genome (AADDG) and cultivated cotton (AADD). The result is a population of individuals in which chromosomes, rather than genes, segregate.

The F2 population is based on progeny derived from the F1 hybrid of *G. nelsonii* x *G. australe*. Initially a subsample of F2 progeny derived from the *G. nelsonii x G. australe* F1 was evaluated using isozyme markers. The isozyme markers confirmed that alleles were segregating among the F2 plants. An F2 population of 101 F2 progeny was sown and 94 individuals were selected as the basis of the genetic mapping population.

The BC1 aneuploid population was derived from germplasm stocks developed under CSP47C. An AADDGG hexaploid, a plant that contains the full genetic complement of *G. australe* (GG) coupled with the full genetic complement of cultivated cotton (AADD), was crossed to cultivated cotton to produce a pentaploid individual (AADDG) with a single copy of each of the 13 *G. australe* chromosomes (G) and two copies of the each of the of cultivated cotton chromosomes (AADD). Crossing this pentaploid individual to cultivated cotton produced 18 aneuploid BC1 individuals among which the 13 *G. australe* chromosomes segregated without recombining. Ten aneuploid BC1 individuals were backcrossed to cultivated cotton to produce 110 BC2 aneuploid individuals.

Extract and test DNA of germplasm stocks and mapping population

Total DNA was extracted from 18 BC1 aneuploid progeny, 110 BC2 aneuploid progeny, 94 *G. nelsonii* x *G. australe* F2 progeny, and all the relevant parental lines. Testing demonstrated that the DNAs were high quality and suitable for PCR based marker analyses, such as AFLPs.

Screen molecular markers against parents of mapping population to identify polymorphisms

Subsequent to development of the two mapping populations and the extraction of high quality DNA, the parental lines were screened for heritable polymorphisms using 31 AFLP primer combinations. As Fig. 2 illustrates, this quickly and efficiently differentiates between markers that are: **A)** common to all the *Gossypium* species used in this study; **B)** unique to cultivated cottons (*G. hirsutum*); **C)** heterozygous in one of the parental lines (P1 & P2); **D)** derived from *G. australe* and segregating among the BC1 progeny and the F2 mapping progeny; **E)** unique to *G. nelsonii* and segregating in the F2 mapping population but absent from the synthetic hybrid and the BC1 individuals; **F)** segregating among the BC1 progeny but shared by the parents of the mapping population and unmappable. Class D markers provide the most useful data because they are segregating in both populations and permit comparisons between the populations, but class E and F markers are also useful because they increase the resolution in at least one of the base maps.

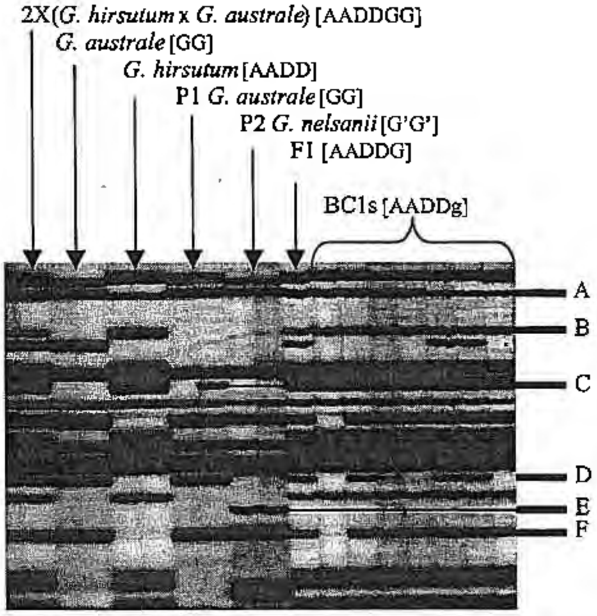


Figure Two. Partial AFLP profile of a synthetic 2X(*G. hirsutum* x *G. australe*) hexaploid, its parents, its F1 hybrid with *G. hirsutum*, and its BC1 progeny (*G.hirstum* was the recurrent parnet) and the *G. australe* and *G. nelsonii* accessions used to synthesize an F2 mapping population. Amplified using primer pairs *Eco*R1-AAG and *Mse*I-MCAG.

Mapping of polymorphisms

Using 31 AFLP primer combinations, 776 *G. australe* alleles were resolved using 31 AFLP primer combinations in the three segregating populations (F2, BC1 aneuploid, BC2 aneuploid): 599 of these alleles segregated in the BC1 and BC2 aneuploid populations, 449 alleles segregated in the F2 population, and 272 alleles segregated in both populations.

Analysis of segregation data

Analysis of the segregation data from the BC1 and BC2 populations is complete. The analysis of the F2 segregation data is nearly complete. The initial maps are complete and all putative double crossovers are being checked manually to eliminate mis-scorings and data transcription errors. Following this, final map orders will be determined.

Continue development of intergenomic germplasm stocks

This is an ongoing process. The current status of a number of stocks of immediate interest are described below:

Backcrossing to *G. hirsutum* of two synthetic hexaploids lines obtained from Belgium continues. An initial pentaploid hybrid between GOS-5271 [2X(*G. hirsutum* x *G. sturtianum*)] and *G. hirsutum* led to the development of 11 BC1 plants. Of these three have produced self-pollinated seed. Ten of these BC1 individuals have been successfully backcrossed to *G. hirsutum*. Two of the BC_2 are self-fertile and are producing enough seed for screening. Because *G. sturtianum* is a possible source of *Fusarium* resistance, 180 BC2 and BC1F2 progeny have been grown out in Canberra. DNA has been extracted from all these plants for a genetic analysis similar to that described here for *G. australe*. The plants have been sent to CSIRO Cotton Research Unit (Myall Vale) for increase. The progeny of these plants will be screened for *Fusarium* resistance.

From the initial pentaploid cross between the synthetic AADDGG hybrid (*G. hirsutum* x *G. australe* hybrid; GOS-5269) and *G. hirsutum*, 18 BC1 plants were obtained. One hundred ten BC2 progeny of 10 of the BC1 individuals were included in the genetic analyses described below. Chromosomes carrying genes conditioning brown lint color and pink flower color have been identified. These lines will be backcrossed further in the hope of obtaining recombinants that carry the genes for these two traits. We also hope to develop monosomic addition lines that carry one of each of the *G. australe* chromosomes. These could provide useful genetic stocks for future research.

Table 1: Synthetic polyploids generated that incorporate wild Australian *Gossypium* species.

Female	Male	Ploidy	Acc #
arboreum	*bickii*	4N	Hyb-776
arboreum	*bickii*	4N	Hyb-792
hirsutum	*anapoides*	6N	Hyb-770
hirsutum	*sturtianum*	6N	Hyb-771
hirsutum	*sturtianum*	6N	Hyb-772
hirsutum	*sturtianum*	6N	Hyb-773
hirsutum	*sturtianum*	6N	Hyb-774
hirsutum	*robinsonii*	6N	Hyb-775
hirsutum	*axiguum*	6N	Hyb-777
hirsutum	*sturtianum*	6N	Hyb-788
hirsutum	*sturtianum*	6N	Hyb-789
hirsutum	*sturtianum*	6N	Hyb-794
hirsutum	*anapoides*	6N	Hyb-795
hirsutum	*sturtianum*	6N	Hyb-793

During the course of this grant, we also confirmed the recovery of 14 new synthetic polyploids (Table 1). The initial priority was to generate sufficient seed numbers to insure that these new lines will not be lost. Following this, a small proportion of seed will be tested from each stock to ascertain the transmission and expression of the "gossypol-free" trait in a *G. hirsutum* background. Four of the synthetic hybrids incorporating the *G. sturtianum* genome will be sent to the CSIRO Cotton Research Unit (Myall Vale) to be tested for *Fusarium* resistance.

In the course of developing synthetic polyploids, we exploited the low frequency of unreduced gamete production in some *G. hirsutum* X *Gossypium* K genome species triploids. By doing so, we generated nine pentaploid derivatives that are the genetic equivalent of BC1s. Five of these have successfully backcrossed to *G. hirsutum*, yielding BC2 seed. These lines will be used as experimental lines in CSP120C.

In the past three years, we generated a number of hybrids between one parent that had 'gossypol-free' seeds and one that did not. One of these hybrid lines was fertile. However, we were unsuccessful in generating F2 progeny. Some BC1 seed was obtained, but with considerable effort and of insufficient numbers for genetic analyses. This strategy was discontinued. An alternate strategy is under consideration. One considerable limitation in trying to ascertain the genetic control of gossypol deposition in the seeds is that the most reliable test (HPLC) is destructive and destroys the desired individual in the process of analyzing it. Non-destructive tests (visual inspection) are not reliable. However, our ability to infer the genetic composition of aneuploid hybrids may permit us to use a modified form of progeny testing to identify chromosomes carrying genes of major effect. In essence we could destructively and reliably assay the progeny of plants of known genomic constitution. The results would then translate back to the parental line. We are currently searching for a skilled analytical chemist who could perform the chemical analyses.

3. Detail the methodology and a justification for the methodology used.

The primary goal of this project was to construct a molecular genetic map of the *G. australe* genome so that we could identify chromosome specific markers for each of the *G. australe* chromosomes. Once this was done, these chromosome specific markers would then be used to track the inheritance of the *G. australe* chromosomes in the BC aneuploids and identify recombinant individuals. Thus it was critical that we identify a cost-effective and high-throughput marker strategy. A variety of molecular marker systems are now available, e.g., SSRs (microsatellites), AFLPs (amplified fragment length polymorphism), RFLPs (restriction fragment length polymorphism). They all have advantages and limitations. One of the key considerations for building genetic linkage maps is resolving a sufficient number of loci to fully resolve all the relationships among the loci. Thus one wants to resolve as many loci as possible in the shortest amount of time. For this purpose, AFLPs were clearly the most efficient system. On a single gel, an average of 25 loci for 60 individuals can be resolved and the total number of loci resolvable is essentially unlimited. In contrast, the maximum number of SSR loci one can realistically resolve on a single gel of 60 individuals is nine, and there is functional limit to the number of potential loci (200-300). Like AFLPs the total number of RFLP loci that can be resolved is essentially unlimited, but the number of resolvable loci on a single gel of 60 samples would average no more than two. Furthermore, RFLPs and SSRs also require 3-6 months of marker development. RFLPs and SSRs are specific to each plant group studied and hence must be redeveloped for new situations. We could have taken advantage of some of the SSRs and RFLPs already being used for other *Gossypium* species, but the amount of time saved would have been minimal. AFLP detection methods are directly applicable to all plant species without redevelopment.

The limitation of AFLPs is that they are dominant markers and the allele of one parent is invisible. The advantage of RFLPs and SSRs is that they are codominant markers and the alleles of both parents are detectable. In this application this limitation was not of concern. The aneuploid BC individuals only have one copy of each of the chromosomes and there is only one allele per locus and in these situations dominant markers carry the same amount of information as codominant markers. For the F2 population, there are two alleles per locus and a codominant system would have been preferred. However, because we are only interested in the alleles of one of the parents, resolving the alternate alleles was not necessary. The need to resolve as many alleles as efficiently as possible was the primary criterion and AFLPs were clearly the most effective option.

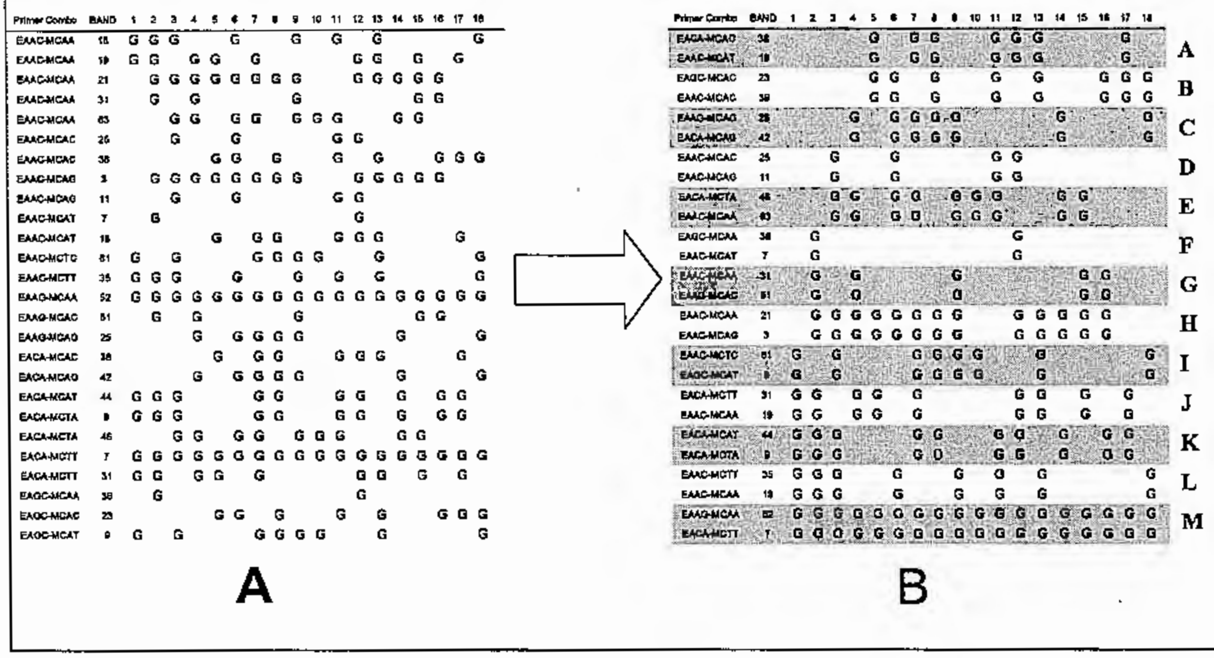
4. Detail results including the statistical analysis of results

- *Using 31 AFLP primer combinations, 599* G. australe *alleles were resolved among 18 BC1 and 110 BC2 aneuploid progeny. Genetic analyses of these data were used to 1) identify the chromosomal location of the* G. australe *alleles, 2) determine the number and the identity of the* G. australe *alleles in the BC1 and BC2 individuals, and 3) determine the location of pink flower color and brown lint color.*

1. The first stage of the development of the genetic maps was to analyze the distribution of the *G. australe*-specific AFLP loci among the 18 BC1 individuals. Because the *G. australe* chromosomes in the BC1 individuals lack partners with which they can pair during meiosis, our expectation is that they would be inherited as unrecombined blocks of loci. This expectation was tested by a genetic analysis designed to identify loci that consistently occur in the same individuals. Figure 3 illustrates this process.

   Of the 599 *G. australe*-specific AFLP alleles assayed, 523 (87%) fell into 1 of 13 unrecombined suites of loci, equal to the number *G. australe* chromosomes (Table 2). Of the remaining 76 alleles, 28 (5%) proved to be two comigrating loci and had to be deleted from the data matrix, and 48 (8%) had eccentric distributions. These latter alleles are the most interesting because they represent putative recombination events. This demonstrated that AFLP alleles from each of the *G. australe* chromosomes were represented in a roughly equivalent manner in the data set (29 to 57).

Figure 3. Genetic analysis of 26 (of 599) *G. australe*-specific AFLP alleles among 18 BC1 aneuploid progeny reveals 13 suites of chromosome specific alleles. A) Raw data matrix. B) Genetic analysis of the *G. australe*-specific AFLPs reveals 13 major suites of unrecombined blocks of alleles that correspond to chromosomes. The letter designations to the right indicate the arbitrarily designations assigned to each chromosome.



Primer Combo	BAND	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
EAAC-MCAA	15	G	G	G			G			G		G		G					G
EAAC-MCAA	19	G	G		G	G		G					G	G		G		G	
EAAC-MCAA	21		G	G	G	G	G	G	G	G			G	G	G	G	G		
EAAC-MCAA	31		G		G					G						G	G		
EAAC-MCAA	63			G	G		G	G		G	G	G			G	G			
EAAC-MCAC	26			G			G					G	G						
EAAC-MCAC	38					G	G		G			G		G			G	G	G
EAAC-MCAG	3		G	G	G	G	G	G	G	G			G	G	G	G	G		
EAAC-MCAG	11			G			G					G	G						
EAAC-MCAT	7		G										G						
EAAC-MCAT	18					G		G	G			G	G	G				G	
EAAC-MCTC	61	G		G				G	G	G	G			G					G
EAAC-MCTT	35	G	G	G			G			G		G		G					G
EAAG-MCAA	52	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
EAAG-MCAC	51		G		G					G						G	G		
EAAG-MCAG	25				G		G	G	G	G					G				G
EACA-MCAC	38					G		G	G			G	G	G				G	
EACA-MCAG	42				G		G	G	G	G					G				G
EACA-MCAT	44	G	G	G				G	G			G	G		G		G	G	
EACA-MCTA	9	G	G	G				G	G			G	G		G		G	G	
EACA-MCTA	46			G	G		G	G		G	G	G			G	G			
EACA-MCTT	7	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
EACA-MCTT	31	G	G		G	G		G					G	G		G		G	
EAGC-MCAA	38		G										G						
EAGC-MCAC	23					G	G		G			G		G			G	G	G
EAGC-MCAT	9	G		G				G	G	G	G			G					G

A

Primer Combo	BAND	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
EACA-MCAC	38					G		G	G			G	G	G				G		A
EAAC-MCAT	18					G		G	G			G	G	G				G		
EAGC-MCAC	23					G	G		G			G		G			G	G	G	B
EAAC-MCAC	38					G	G		G			G		G			G	G	G	
EAAG-MCAG	25				G		G	G	G	G					G				G	C
EACA-MCAG	42				G		G	G	G	G					G				G	
EAAC-MCAC	26			G			G					G	G							D
EAAC-MCAG	11			G			G					G	G							
EACA-MCTA	46			G	G		G	G		G	G	G			G	G				E
EAAC-MCAA	63			G	G		G	G		G	G	G			G	G				
EAGC-MCAA	38		G										G							F
EAAC-MCAT	7		G										G							
EAAC-MCAA	31		G		G					G						G	G			G
EAAG-MCAC	51		G		G					G						G	G			
EAAC-MCAA	21		G	G	G	G	G	G	G	G			G	G	G	G	G			H
EAAC-MCAG	3		G	G	G	G	G	G	G	G			G	G	G	G	G			
EAAC-MCTC	61	G		G				G	G	G	G			G					G	I
EAGC-MCAT	9	G		G				G	G	G	G			G					G	
EACA-MCTT	31	G	G		G	G		G					G	G		G		G		J
EAAC-MCAA	19	G	G		G	G		G					G	G		G		G		
EACA-MCAT	44	G	G	G				G	G			G	G		G		G	G		K
EACA-MCTA	9	G	G	G				G	G			G	G		G		G	G		
EAAC-MCTT	35	G	G	G			G			G		G		G					G	L
EAAC-MCAA	15	G	G	G			G			G		G		G					G	
EAAG-MCAA	52	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	M
EACA-MCTT	7	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	

B

2. The recognition of these 13 suites of chromosome-specific AFLP alleles revealed the number and the identity of the *G. australe* chromosomes in the 18 BC1 (Table 3) and 110 BC2 individuals (data not shown). Of the 234 *G. australe* chromosomes available, 108 (46%) were transmitted through the first meiosis. Individual BC1 plants had three to eight (average 6) of the *G. australe* chromosomes. Thus in this first generation, by both measures, just under half of the *G. australe* chromosomes were retained. The results for the 110 BC2 individuals are congruent. Of the 685 chromosomes available for transmission, 299 (45%) were retained. The number of chromosomes per individual BC1 ranged from one to six, with an average of 2.7. Again, by both measures just under half of the available chromosomes were transmitted through the second meiosis.

   Among the 18 BC1 progeny, the number of chromosomes per individual was roughly equivalent, but the frequency of transmission for the respective *G. australe* chromosomes was more variable, ranging from 2 to all 18 individuals. There appeared to be a compensatory response among the 10 BC2 families (data not shown). The number of chromosomes per backcross individual was more variable, but the frequency of chromosome transmission was roughly equivalent for all the chromosomes. Thus, with one notable exception (see below) there are no outstanding biases in chromosome transmission. Chromosomes A through L, are found in 6 to 34% of the BC2 progeny. No chromosomes have been selectively eliminated in the first two generations.

   The one notable exception is chromosome M. As demonstrated in Table 3, all 18 BC1 progeny inherited chromosome M. While this was unlikely, it was not overtly improbable. Among the BC2 progeny, we expected that chromosome M would be retained by some but not all of the progeny. This was the pattern for the other 12 *G. australe* chromosomes. Surprisingly this was not the case; all 110 BC2 progeny inherited chromosome M. Explanations attributing this observation to a variety of methodical artifacts are unsatisfactory, but so is the proposition that this occurred simply by chance. While a satisfactory biological explanation is still wanting, the hypothesis that chromosome M is required for the mitotic stability or the physiological competence of *G. australe* x *G. hirsutum* aneuploids is under consideration. If this is true, it represents a critical component of our understanding of these aneuploid hybrids and how they can be used in cotton breeding. Further studies are planned to explore this result further.

Table 2. Number of AFLP alleles per *G. australe* chromosome.

CHROMOSOME	NO. OF MARKERS
A	55
B	44
C	36
D	55
E	35
F	40
G	32
H	57
I	34
J	29
K	38
L	36
M	32
Recombinants?	48
2 alleles	28

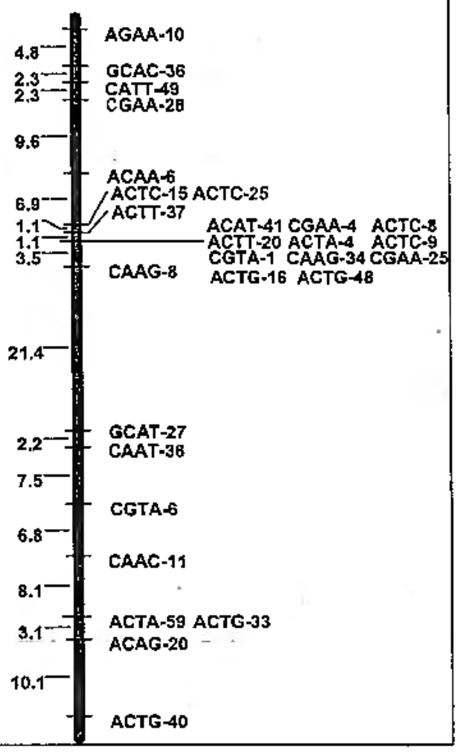
Table 3. Distribution of *G. australe* chromosomes among 18 BC1 aneuploid progeny.

Chromosome	668	669	670	671	681	682	699	700	701	702	703	704	705	706	707	708	709	754	No. of BC1s
A					X		X	X			X	X	X				X		7
B					X	X		X			X		X			X	X	X	8
C				X		X	X	X	X					X				X	7
D			X			X					X	X							4
E			X	X		X	X		X	X	X			X	X				9
F		X										X							2
G		X		X					X						X	X			5
H		X	X	X	X	X	X	X	X			X	X	X	X	X			13
I	X		X				X	X	X	X			X					X	8
J	X	X		X	X		X					X	X		X		X		9
K	X	X	X				X	X			X	X		X		X	X		10
L	X	X	X			X			X		X		X					X	8
M	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	18
No. of Chromosomes	5	7	7	6	5	7	8	7	7	3	7	7	7	5	5	5	5	5	

3. Knowing the chromosome complement of the BC1 and BC2 progeny allowed us to identify the chromosomal location of conditioning morphological traits under the control of single dominantly expressed genes. The two morphological traits (pink flower color, brown lint) considered here are not of high priority for the Australian cotton industry, although they be useful, but our ability to identify the chromosomes is instructive. The ability to do so was unexpected and this strategy will be applied to traits of higher priority, e.g., *Fusarium* resistance and gossypol-free seeds. In essence this analysis is simple, and is based on the correlatinn between expression of the trait and the presence of specific chromosomes. Of the 18 BC1 individuals, only plants 669 and 704 retain the pink flower color of *G. australe*, and these are the only individuals that carry chromosome F (Table 3). The implication that chromosome F carries the relevant is reinforced by the observation that the only BC2 individuals with pink flowers are progeny of 669 and 704 and still carry chromosome F. Similarly the only individuals with brown lint (681, 699, 700, 704, 705, and 709: 703 has not produced seed and its lint color is unknown) are the only individuals that inherited chromosome A (Table 3). Again the BC2 data are fully congruent.

- *Using the same 31 AFLP primer combinations described above, 449* G. australe *alleles were mapped in a population of 94 F2 progeny derived from the F1 hybrid between* G. nelsonii *and* G. australe. *Genetic analysis revealed that these 1) 449 mapped to 269 distinct genomic regions, 2) 18% of the loci mapping to three distinct genomic regions are under selection, 3) the 269 distinct genomic regions represent 17 linkage groups (13 linkage groups with 26+loci and 4 ancillary linkage groups with less than 5 loci).*

Figure 4: Genetic linkage map of *G. australe* chromosome F.a1



1. The first step of the analysis was to reduce the size of the data matrix; 449 loci are computationally unwieldy. Because many loci that originate from recombinationally inert regions of the genome cannot be differentiated statistically in a population of 94 individuals, it is computationally more efficient to pool them into a single locus. This demonstrated that 449 loci mapped to 269 recombinationally distinct genomic regions: 229 of these genomic regions were represented by a single allele, 27 were represented by small clusters (5 or less) of alleles, and 13 were represented by large allele clusters (5-20). This last class of alleles is particularly interesting because they may arise from the centromeric regions of the chromosomes (see cluster of 11 loci between *ACTT-37* and *CAAG-8* in Fig. 4).

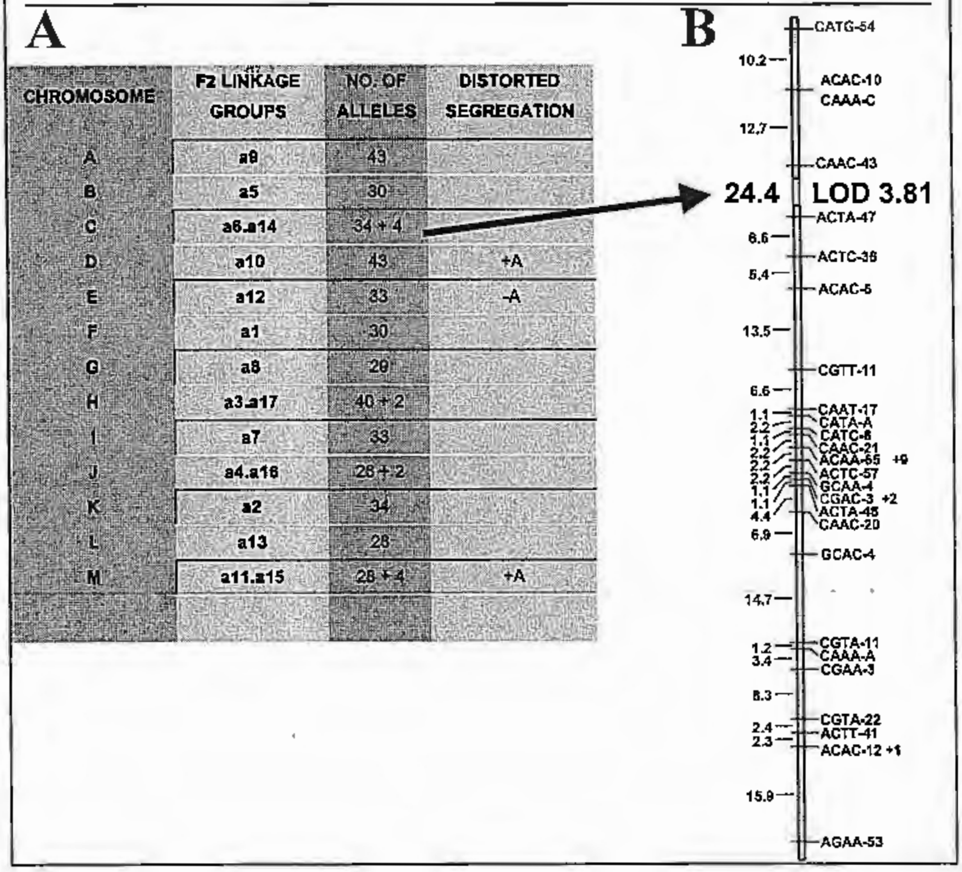
2. The second analysis determined how many of the loci deviated significantly ($P<0.05$) from Mendelian expectations, i.e, distorted segregation. In a segregating population, the expectation is that about 5% of alleles would exhibit distorted segregation as a function of statistical "noise", however, the observation that 18% of the loci exhibited distorted segregation suggests some of the loci were truly under selection. That these loci mapped to three distinct genomic regions reinforces this supposition. While nnderstanding the biological reasons for this result is not of high priority, the knowledge that these loci were under selection was important in the final stage of the analysis.

3. The final stage of this analysis was a genetic analysis that allowed us to infer the respective physical relationships among the loci within each linkage group. This in turn provides a linear map of the *G. australe* chromosomes (e.g., Fig. 4). The result is a highly resolved genetic linkage map of the *G. australe* genome. Seventeen linkage groups were identified: 13 with 26 or more loci and four ancillary linkage groups with less than five loci.

- *The final step of the overall analysis is the comparison of the segregation data from the BC1 and BC2 progeny to the genetic linkage map of the* G. australe *genome. These comparison are coinformative, allowing us to 1) more fully resolve the genetic linkage map of the* G. australe *genome, and 2) identify recombinant* G. australe *chromosomes in the BC1 and BC2 progeny.*

1. The genetic analysis of the *G. nelsonii* x *G. australe* F2 population revealed 13 large linkage groups (26+ loci) and four small ancillary groups (less then five loci). Because *G. australe* has 13 chromosomes a fully resolved genetic linkage map will converge on 13 linkage groups. Resolution of genetic linkage maps is dependent on the number of markers; up to a certain threshold level, the level of resolution is proportional to the number of markers. To more fully resolve the genetic linkage map, we would have to continue adding segregation data from additional loci. However, a more efficient way to do this is to use the segregation data from the BC1 and BC2 populations. Because the chromosomes in these populations are inherited as unrecombined blocks, linkage among loci is not dependent on marker density. These comparisons allowed us to fully resolve the genetic linkage map to 13 linkage groups corresponding to the 13 chromosomes of *G. australe*. Figure 5 illustrates one example and details the final map comparisons.

Figure 5. Comparison of segregation data for BC1 aod BC2 populations and the *G. nelsonii* x *G. australe* F2 population. A) Correspondance of the chromosome groups recognized in the BC1 and BC2 populations and the linkage groups identified in the genetic analysis of the *G. nelsonii* x *G. australe* F2 population. B) Linkage C.a6.a14, illustrating how two F2 linkage groups can be combined based on the information from the BC1 and BC2 individuals.

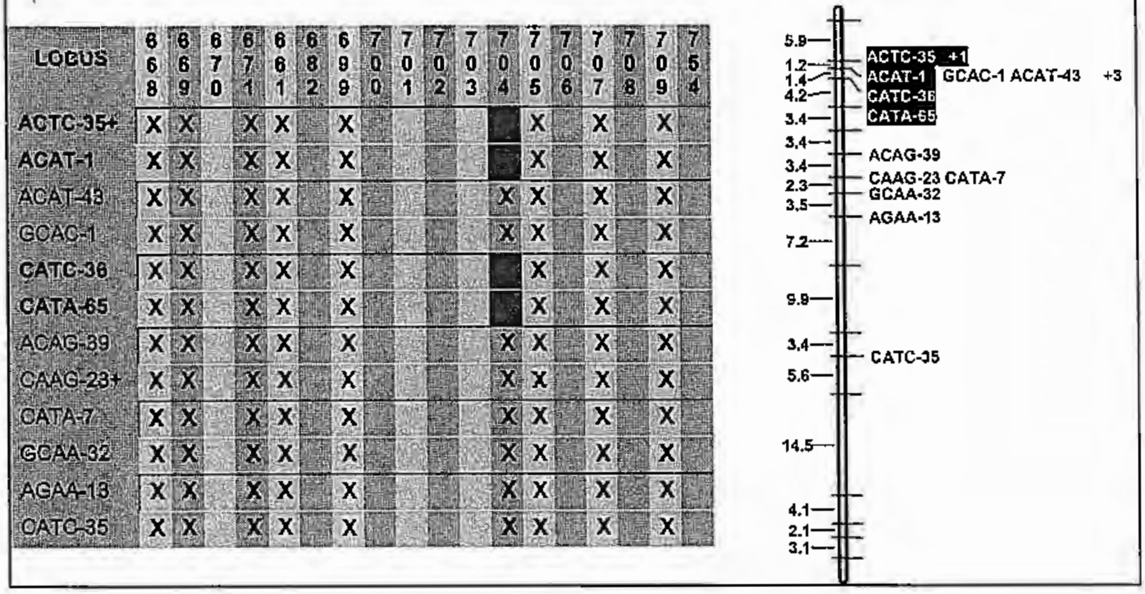


A

CHROMOSOME	F2 LINKAGE GROUPS	NO. OF ALLELES	DISTORTED SEGREGATION
A	a9	43	
B	a5	30	
C	a6.a14	34 + 4	
D	a10	43	+A
E	a12	33	-A
F	a1	30	
G	a8	29	
H	a3.a17	40 + 2	
I	a7	33	
J	a4.a16	28 + 2	
K	a2	34	
L	a13	28	
M	a11.a15	28 + 4	+A

2. The final analysis comprises checking the genomic location of the putative recombinants identified in Table 2. These markers have inheritance patterns that are inconsistent with our expectation that chromosomes in the BC1 and BC2 progeny would be lost or retained as unrecombined blocks of loci. Figure 6 illustrates one example. The ability to do this is highly advantageous. This allows us to determine how frequently and which of the *G. australe* chromosomes have been involved in recombination events. The comparisons revealed that 6 of the 13 *G. australe* chromosomes have been involved in 14 putative recombination events.

Figure 6. Loss of four *G. australe* AFLP alleles from chromosome J.a4. in BC1 individual 704 A) Distribution of mapped AFLP alleles among the 18 BC2 indviduals. Note absence of alleles ACTX-35+, ACAT-1, CATC-36, and CATA-65 from individual 704. B) Genetic linkage map of *G. australe* chromosome J.a4. The genomic location of the four alleles missing from BC1 indvidual 704 are higlighted.



LOCUS	668	669	670	671	661	682	699	700	701	702	703	704	705	706	707	708	709	754
ACTC-35+	X	X		X	X		X					■	X		X		X	
ACAT-1	X	X		X	X		X					■	X		X		X	
ACAT-43	X	X		X	X		X					X	X		X		X	
GCAC-1	X	X		X	X		X					X	X		X		X	
CATC-36	X	X		X	X		X					■	X		X		X	
CATA-65	X	X		X	X		X					■	X		X		X	
ACAG-39	X	X		X	X		X					X	X		X		X	
CAAG-23+	X	X		X	X		X					X	X		X		X	
CATA-7	X	X		X	X		X					X	X		X		X	
GCAA-32	X	X		X	X		X					X	X		X		X	
AGAA-13	X	X		X	X		X					X	X		X		X	
CATC-35	X	X		X	X		X					X	X		X		X	

5. Discuss the results, and include an analysis of research outcomes compared with objectives.

This research program sought to 1) continue the development of the interspecific germplasm initiated under CSP47C, 2) Develop a genetic linkage map of the Australian G genome, and 3) to develop chromosome specific molecular markers to trace the transfer of genetic material from the indigenous Australian Gossypium *species to cultivated cottons.*

1. The goal of continuing development of interspecific germplasm that incorporates chromosomes of the indigenous Australian *Gossypium* species into a cultivated cotton background is to make use of valuable traits in these native species. Most notable are the lack of gossypol in the seed (a trait only found in Australian *Gossypium* species) and potential resistance to the indigenous *Fusarium* wilt that is a serious disease in eastern cotton growing districts. Because the Australian *Gossypium* species do not readily form fertile hybrids with cultivated cottons this is a difficult task. However, a carefully planned crossing program and diligent effort enabled us to breed various backcross lines that proved valuable in this project. From the BC2 aneuploid lines that have been used in this study, BC3 and BC2F2 populations that carry genes for pink flower color and brown lint are being developed. Knowing which of the chromosomes carries the relevant gene in each case allows us to focus on a few relevant individuals, which is far more efficient than a "shotgun" approach.

   Most importantly, a number of germplasm stocks that we expect will carry *Fusarium* resistance are now available. The *Fusarium* strains that are currently found in the eastern cotton districts are unique to Australia and are likely to have arisen here. Thus it is probable that that indigenous Australian *Gossypium* species have had previous exposure to these *Fusarium* strains. Preliminary testing by the QLD DPI suggests that some of the K genome species and *G. sturtianum* are resistant. As a result, 180 BC2 aneuploids individuals that carry *G. sturtianum* chromosomes have been sent the CSIRO Cotton Research Unit in Myall Vale (NSW) for increase so that they can included in glasshouse and field trials for *Fusarium* resistance. We have also developed a BC1 population of aneuploids that carry chromosomes from a K genome species. The K genome parent in this instance has not yet been tested for *Fusarium* resistance, but if it does prove to be resistant, these lines will also be sent to the CSIRO Cotton Research Unit for increase for glasshouse and field trials. While much of this germplasm development will be more fully realized under the new CRDC grants CSP120C and CSP113C, it illustrates the value of judicious long-term investment in germplasm development.

Although the focus of development in the new CRDC grant CSP120C will switch to germplasm lines with possible *Fusarium* resistance, we will continue development of the lines incorporating the gossypol-free seed trait developed during this grant. This characteristic presents a challenge because screening for expression of the desired trait—visible only in seed tissue—can destroy the seed. Non-destructive tests are unreliable. However, we now have germplasm that may allow us to circumvent this problem. Among the BC2 population are individuals that carry only one of the *G. australe* chromosomes (in addition to the ever-present chromosome M), and at least one of these lines is available for each of the 13 *G. australe* chromosomes. This should allow us to examine the seed of these individuals, and attribute differences in seed gossypol expression patterns to specific chromosomes. We plan to continue this work through the life of the next grant.

2. The development of a genetic linkage map for the Australian G genome was a high priority goal of this project. Genetic linkage maps are essentially roadmaps to genomes. In this instance, we wanted to track the inheritance of the *G. australe* chromosomes in the BC aneuploid lines and to characterize their interactions with the *G. hirsutum* chromosomes. To do this we needed chromosome specific markers. We needed to know exactly from which of the 13 chromosomes each of the markers belonged. We initially proposed using a F2 population to construct a genetic linkage map. Traditionally the level of resolution in a genetic linkage map is proportional to the number of loci mapped, and complete resolution (one linkage group per chromosome) requires very large numbers of loci. However, we realized during the course of work that rather than relying entirely on the F2 population to elucidate the physical relationships among the loci, we could also use the BC1 individuals to determine the chromosomal location of all markers regardless of the number of markers used. Thus we were able to more efficiently develop a fully resolved genetic linkage. Furthermore the linkage relationships are supported by data from two independent populations rather than just one.

   The other key and unforeseen advantage is that we can now determine the chromosomal location of markers that are not segregating in the F2 population. In an F2 population, one can only determine the genomic location of loci for which the two parents differ genetically. This has been a major impediment in mapping the cultivated cotton genome. In practice, the parents have two distinct alleles for only a subset of the loci. Distantly related parental lines will have greater numbers of allelic differences and therefore a higher proportion of mappable loci. Unfortunately, as the distance between the parental lines increases, the frequency of confounding artifacts (e.g., segregation distortion) also increases, decreasing the statistical reliability of the data. However, using the whole chromosome segregation among only 18 BC1 individuals, we could determine the chromosomal location of loci that could not be mapped in the F2 population. The limitation is that linear order of these non-mappable could not be determined, but with no additional effort, we vastly increased the information content of the data.

3. The impetus for this project was the difficulties experienced in transferring genes from the indigenous Australian *Gossypium* species. Having developed these BC1 and BC2 aneuploid populations, we needed a means of tracking the fate of the *G. australe* chromosomes placed in proximity to the chromosomes of the cultivated cottons. We wanted to determine how quickly the *G. australe* chromosomes are lost, to what extent they exchange material with the chromosomes of cultivated cotton, and to identify recombinants. Using the genetic maps described in #2, we were

able to determine that roughly 45% of the *G. australe* chromosomes are lost every generation. We also established that, with the exception of chromosome M, there are no strong biases in chromosome transmission frequencies. In the process, fourteen recombination events involving six of the 3 *G. australe* chromosomes were identified. This provides a previously unavailable view into the interaction between chromosomes of two different species. Our ability to identify recombinants with molecular markers means that when we identify recombinants involving agronomically useful traits, pre-identified molecular markers for marker assisted breeding will accompany them.

Overall, the results demonstrate that we developed a more efficient means of incorporating agronomically valuable traits from the indigenous Gossypium *species into cultivated cotton. Using molecular genetic maps, we can monitor the inheritance of the donor chromosomes, identify chromosomes carrying the genes of interest (for traits under simple genetic control), identify recombinants, and will have molecular markers available for marker assisted breeding.*

6. Provide an assessment of the likely impact of the results and conclusions of the research project for the cotton industry. Where possible include a statement of the costs and potential benefits to the Australian cotton industry and future research needs.

The results of this project will benefit the cotton industry over the next ten years in a number of direct and indirect ways. The major product so far is cotton-breeding lines incorporating the chromosomes of the indigenous Australian *Gossypium* species. With these breeding lines, we hope to incorporate agronomically valuable traits (e.g., *Fusarium* resistance and gossypol-free seeds) into future elite cotton cultivars. Material with potential *Fusarium* resistance is already being increased for field and glasshouse trials. If this germplasm proves promising, the molecular markers and the strategies of analysis developed here will be important tools in developing this germplasm into commercially viable cotton cultivars. Whether the *Fusarium* resistance in the indigenous *Gossypium* species will be expressed in the cultivated cotton background, how many genes control the *Fusarium* resistance of the native species, and whether the gene(s) involved can be transferred to the chromosomes of cultivated cottons are key research questions to be addressed in the current CRDC grants, CSP120C. Addressing these questions is just one facet of CSIRO's current efforts mitigate the effects of *Fusarium* induced wilt in the Australian Cotton Industry. The research methods, molecular markers, and germplasm that were realized during CSP85C will be invaluable tools in this pursuit.

7. Describe the project technology (e.g. commercially significant developments, patents applied for or granted, licenses, etc).

Not applicable

8. Provide a technical summary of any other information developed as a part of the research project. Include discoveries in methodology, equipment design, etc.

Not applicable

9. State the recommendations on the activities or other steps that may be taken to further develop, disseminate, or to exploit the project technology.

Meeting the objectives of this grant generated considerable experience in the application of molecular markers to difficult breeding objectives. Under CRDC grant CSP120C, the use of molecular markers in cotton cultivar development will continue. Many of the skills and resources developed for these projects can be applied to a wide range of objectives. Having developed this resource base, it would be beneficial to train young cotton researchers. As a first step toward this, a CRC funded post-graduate student, Augusto Becerra, has been appointed as a post-doctoral fellow under CRCD grant CSP120C. That he was the preferred candidate illustrates the utility of investing current experience in promising post-grads. To this end, under we will explore opportunities for including post-graduate students in CSP120C.

While CRDC grant CSP120C will continue the development of new germplasm with valuable agronomic traits and the molecular marker methods needed to do this efficiently, there is another way in which these resources and skills can expand this resource. This project illustrates the power of molecular markers to track genetic events in interspecific germplasm, but we are limited in that we can only infer the physical chromosome interactions that are responsible for the events observed. The genetic data described here would be more powerful if they could be coupled with cytogenetic (direct physical) observations of the chromosome interactions. These comparisons provide insight into the physical processes responsible for exchanges of genes between chromosomes of cultivated cotton and those of native Australian species. Understanding these physical processes may lead to ways in which they could be manipulated to benefit breeding goals. The limited exchange of genes is and will continue to be the primary barrier to transferring valuable genes from the indigenous Australian *Gossypium* species to cultivated cottons. Cotton cytogenetic expertise does not exist within the Australian cotton research community. Developing these skills in Australia would have immediate benefit to the development of interspecific germplasm and in the longer term will be necessary as the Australian Cotton Research community expands its involvement in international cotton genomics research. To this end, a new grant application will be put forward in the next funding round for a post-doctoral fellowship that would focus on the physical chromosome interactions between cultivated cottons and those of the indigenous *Gossypium* species.

10. List the publications arising from the research project.

Brubaker, C. L., and A. H. D. Brown. In preparation. A genetic linkage map of the Australian G genome. For *Genome.*

Brubaker, C. L., and A. H. D. Brown. In preparation. Inheritance of *Gossypium australe* chromosomes and homoeologous recombination in alien chromosome addition lines of *Gossypium hirsutum.* For *Genome.*

Brown, A. H. D., and C. L. Brubaker. 2000. Genetics and the conservation and use of Australian wild relatives of crops. *Australian Journal of Botany*48:297-303.

Brubaker, C. L., A. H. D. Brown, J. McD. Stewart, M. J. Kilby, and J. P. Grace. 1999. Production of fertile hybrid germplasm with diploid Australian *Gossypium* species for cotton improvement. *Euphytica* 108:199-213.

Seelanan, T., C. L. Brubaker, J. McD. Stewart, L. A. Craven and J. F. Wendel. 1999. Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Systematic Botany* 24: 183-208

Wendel, J. F., R. L. Small, R. C. Cronn, and C. L. Brubaker. 1999. Genes, jeans, and genomes: Reconstructing the history of cotton. In L.W.D. van Raamsdonk and J. C. M. den Nijs [eds], Plant evolution in man-made habitats. Proceedings of the VII^th^ international symposium of the international organization of plant biosystematists, pp. 133-161. Hugo de Vries Laboratory, Amsterdam, The Netherlands.

Part 5 - Plain English Summary

Within Australia are there 17 species of native cotton relatives. These native species are wild plants found across the Australian continent and nowhere else in the world. Although they are distant relatives of the commercial cottons grown for the production, they have characters that could be used to improve commercial cotton cultivars. Preliminary tests identified some species with resistance to the *Fusarium* wilt that is a serious disease threat in many cotton growing regions. Other species lack the toxic gossypol in the seeds that must be removed before the seed protein can be used in animal feeds and before the oil can be used for human consumption.

Transferring the genes responsible for these desirable characteristics from the native species to the cultivated cottons is difficult. Because the native cotton relatives are distant relatives of the cultivated cottons, traditional breeding strategies are not applicable. Hybrids between the native cottons and cultivated cottons can only be made fertile with human intervention and the transfer of genetic information from the chromosomes of native species to those of the cultivated cottons is infrequent. Fortunately, this means that genes from genetically modified cotton cultivars cannot escape into native cotton populations, but it makes it difficult to transfer valuable traits from the native cottons to the commercial cultivars. With diligent effort, we have been successful in creating fertile hybrid plants that combine the wild species with the cultivated cottons. One of these lines combines a wild species with gossypol-free seeds with the cultivated cottons and the other is a hybrid between cultivated cotton and a species that has potential *Fusarium* resistance. These breeding lines provide the opportunity for some gene exchange.

To increase our chances of finding these rare gene exchange events, we developed molecular markers that identify each individual chromosome of the wild species in the hybrid breeding lines. With these markers we constructed a genetic linkage map of the wild cotton species with gossypol-free seeds. Genetic linkage maps are essentially road maps to genomes of plants (and animals). As these hybrid plants progress through the breeding program, the genetic linkage map allows us to track the inheritance of the wild genes, to detect when wild genes have moved onto cotton chromosomes, and to identify the wild chromosomes that carry important genes. With this knowledge, we can more efficiently overcome the difficulties these unique hybrids present. Research to date focused on the native species that have gossypol-free seeds, but over the next three years, the experience and resources will now focus on the species with potential *Fusarium* resistance.