CTA046

FINAL REPORT

TITLE:	**PERFECT MARKERS FOR SUGARCANE MAPPING**
PROJECT NUMBER:	CTA046
RESEARCH ORGANISATION:	CSIRO Plant Industry

PRINCIPAL INVESTIGATOR:	Dr Lynne McIntyre
Phone:	07 3214 2321
Address:	CSIRO Plant Industry 120 Meiers Rd, Indooroopilly, 4068

The research presented in this report has been funded in part by the Sugar Research and Development Corporation.

DISCLAIMER: "The Research Organisation is not a partner, joint venturer, employee or agent of SRDC and has no authority to legally bind SRDC, in any publication of substantive details or results of the Project".

EXECUTIVE SUMMARY

Sugarcane is a complex aneuploid, polyploid, interspecific hybrid. Most breeding traits are complex and molecular markers associated with these traits are a tool that may assist breeders to more efficiently identify sugarcane clones containing these desirable traits. At the start of this project, molecular mapping in sugarcane was relatively new with maps being developed in interspecific crosses or within the selfed progeny of a variety. These maps relied heavily on the use of two major marker types, namely restriction fragment length polymorphisms (RFLPs) and randomly amplified polymorphic DNA (RAPDs), both of which had major disadvantages in their ease and reliability of use. The area of genomics was also very new, and genomics studies of sugarcane were being initiated in Australia, the USA, Brazil and South Africa, generating large amounts of gene information and potential markers for mapping in many traits of interest. The major objectives of this project were, therefore, to compare genomic regions associated with traits in Australian and French sugarcane populations, to demonstrate whether "candidate genes", derived from partial gene sequences (termed expressed sequence tags, ESTs), were a useful strategy for the identification of closely linked markers for traits of interest, and to develop methods other than the RFLP approach to enable more rapid screening of ESTs in mapping populations. Using a variety of approaches, these objectives have been met as described below.

Australian sugarcane maps have now been generated from two crosses involving Australian sugarcane varieties or elite material. Each cross has been used to generate two maps, one for the male and female parent of the cross. The Q1 population is derived from a cross between Q117 and 74C42, while the Q3 population is derived from a cross between Q117 and MQ77-340. The four maps now have approximately 300-450 markers each, using mainly PCR-based markers, namely simple sequence repeat (SSRs) and amplified fragment length polymorphisms (AFLPs). Our policy of using SSRs that have been mapped in the French variety R570, currently the largest published map with approximately 1000 markers, has assisted our comparative mapping attempts between the Australian and French maps. Many of the co-segregation groups identified in the Q1 and Q3 population maps have been tentatively aligned with Linkage Groups in R570 on the basis of shared SSRs. Using this approach, we have shown that the same SSRs are linked to sugar and fibre-related traits in both R570 and Q3 populations. This observation provides additional support to the usefulness of these SSRs in marker-assisted selection. In addition, novel associations were also identified for both Q3 and R570 populations. These marker-trait associations could be due to different parts of the genome being mapped in the two populations, to novel alleles in the two populations, or to fixation of this region in one of the populations. The alignment between the R570 and Q3 population maps was more difficult than anticipated because most markers on the R570 map were AFLPs and because marker sizes had not been recorded for either map. AFLPs are plentiful but are more difficult to score accurately and comparatively. In addition, the low level of genome coverage in both maps (~33% for R570 and <20% for Q3) meant that genome comparisons could

only be made at the homology group level as it was likely that different linkage groups within a homology group had been mapped in the two populations.

Candidate genes are potentially an extremely useful source of markers for marker-assisted selection because they may be the gene responsible for expression of the trait. If so, there will not be any recombination between the candidate gene and the trait and thus they are a perfect indirect selection tool. Our results to date are encouraging. We have identified candidate genes that are associated with disease resistance, sugar-related traits and fibre in both Australian and the French sugarcane mapping populations. Further research is now needed to determine whether the association is coincidental, in which case the candidate gene is still a useful marker, or whether the gene is implicated in the expression of the trait, in which case it could provide valuable information on the underlying genetics, biochemistry and physiology of the trait itself. We have also demonstrated that candidate genes appear to be more efficient at identifying marker-trait associations than random markers such as AFLPs and SSRs. A major limitation to the investigation of the usefulness of candidate genes has been the difficulties associated with RFLP analysis in sugarcane.

In addition to RFLP analysis, several other EST mapping techniques were evaluated, including SSRs within ESTs, real-time PCR and allele-specific primers. All four techniques were successful. Of the four techniques, RFLPs were technically the most difficult but permitted the greatest number of ESTs to be evaluated. EST-derived SSRs were the most user-friendly techniques, but limited to those ESTs containing SSR sequences (approximately 4%). In addition, these SSRs were not very polymorphic within sugarcane but were most useful in inter-specific comparisons. Allele-specific primers were very promising for mapping allelic variants of a small number of specific genes. Real-time PCR was also very promising in detecting dose variation of specific alleles.

This project has successfully met its objectives. It has demonstrated the value of comparative mapping between small and larger sugarcane maps, the usefulness of the candidate gene approach as a source of useful markers in sugarcane, and the variety of candidate gene mapping techniques currently available. It has also demonstrated the difficulties associated with sugarcane’s complex genetic structure. Further work is required to confirm the role of these candidate genes in the targetted traits and to identify and map specific alleles in these genes for marker-assisted selection.

PROJECT BACKGROUND

The complexity of the sugarcane genome necessitates a very large number of markers to obtain reasonable genome coverage. It is an aneuploid, polyploid, interspecific hybrid. Consequently there are large (approximately 9-13) and presumably varying numbers of homologous and homoeologous chromosomes. Given the size of our mapping populations, only single dose markers segregating 1:1 (or 3:1 if present in both parents) can be used. Thus, in order to map each member of a homology group, multiple and linked single-dose markers need to be identified for each member of that chromosome homology group. In practice, it would appear that each map needs to be in excess of 500 markers to obtain the number of linkage groups equivalent to the estimated number of chromosomes. However, even with 1000 markers, the estimated genome coverage is only 33-50% (Hoarau et al, 2002; K. Aitken, pers. comm.), indicating that while each linkage group has been identified, all homologous/ homoeologous regions of the genome that pertain to a targetted trait have not been identified and tagged. Consequently, our QTL analyses are undertaken with only a portion of the genome and only a limited proportion of the genes.

Molecular markers have the potential to greatly assist plant breeders through marker-assisted selection, fingerprinting of varieties, characterisation of germplasm etc. Detailed genetic maps are currently available for most of the major agricultural species. It has recently been found that there is considerable conservation of gene and marker order amongst the grass species. Work is currently underway in several laboratories to align the many genomes of the Poideae grasses using “anchor” probes which have already been mapped in several species and hybridise across a wide range of grass species. Comparative mapping is being used to “fill-in” gaps on specific linkage maps. Of particular interest, however, is recent research in maize and rice demonstrating the use of comparative mapping to predict the location of a gene (pistillate) in rice, based on the known location of similar gene in maize (tassel seed). This approach could be extended in sugarcane to look for genes controlling agronomic traits, such as sugar accumulation, using markers which are linked to known components of sugar accumulation in other species or crosses, such as the sucrose or carbohydrate biosynthetic pathway.

In sugarcane, comparative mapping could provide an opportunity to increase the genome coverage of individual maps, as large numbers of markers are required to obtain even low levels of genome coverage. Recent maps by Hoarau et al (2002) and Aitken et al (pers comm.) have more than 1000 markers on them, but still only have 33-50% of the genome covered. In any other crop, this number of markers would be exceptional. Comparative mapping between sugarcane maps could be used to identify new markers and new regions of linkage groups that are associated with traits and are not covered in all sugarcane maps, as well as to identify common QTLs which are therefore likely to be robust.

OBJECTIVES

1. **To identify and request "candidate genes" in the sucrose and carohydrate metabolism, fibre synthesis and disease resistance pathways. Search EST databases, especially the SRDC/CSIRO sugarcane EST database, for "candidate" genes.**

Achieved:

Approximately 400 sugarcane ESTs were identified as belonging to categories of relevance to the sugar, fibre and disease resistance traits segregating in the French and Australian sugarcane mapping populations. In addition, 55 resistance gene analogues (RGAs) were also identified from sugarcane and tested as candidate genes.

2. **Map "candidate genes" in 2 populations (one from BSES and the other from CIRAD) that are segregating for CCS, fibre and disease resistance.**

Achieved:

Candidate genes for sugar and fibre:
Thirty-two putative candidate genes were screened across 112 progeny of R570 of which twenty-three gave scorable single-dose patterns. These 23 genes generated 139 markers of which 83 were single dose. Sixty-five of the 83 markers have been included in 35 co-segregation groups in the map of R570; the remaining 18 markers are unlinked. These candidate genes include genes from the carbohydrate and fibre biosynthetic pathways.

Candidate genes for disease resistance:
Rust resistance in R570 has previously been reported to be due to a major gene and several RFLP probes were shown to be linked to this gene, including CDSR29, SSCIR194 and TXS1257, with CDSR29 being the closest marker. These 3 RFLP probes were screened across progeny of the Q1 (Q117 x 74C42) mapping population. Both CDSR and SSCIR194 gave scorable single-dose markers (4 and 3, respectively).

In addition, of the 55 RGAs isolated from sugarcane, 34 generated single-dose markers and were screened across the Q1 population, generating 73 markers. The Q1 population has been phenotyped for common rust resistance and Pachymetra resistance, as part of CTA049.

We have also mapped approximately 100, 150, and 150 sugarcane cDNA clones (ESTs) in 3 sorghum mapping populations. Given the complexity of the sugarcane genome, the sorghum map (2x=2n=20 chromosomes) is providing a useful reference point for simplifying our sugarcane maps.

3. **Measure correlation between CCS, sugar yield, fibre and disease resistance and "candidate genes" in order to see if the candidate genes provide "perfect" or better markers for these traits than current marker techniques.**

Achieved:

In all populations tested, candidate genes consistently gave a higher probability of finding an association with a trait, than did anonymous RFLP, SSR or AFLP markers.

In R570, the 65 markers generated by the 23 candidate genes were screened across 112 progeny from the population. Phenotypic information for 4 traits (Brix, height, stalk number and sucker number) was obtained by Dr Jean-Yves Hoarau on Reunion Island. Four markers were consistently associated with a trait over the two years for which data was collected (one for Brix, one for height and two for stalk diameter). These significant markers represent 6% of the candidate gene markers tested. By contrast, Hoarau et al (2002) found 40 markers from a total of 1000 AFLP markers to be associated with these 4 traits, of which only 5 appeared in both years. This represents only 0.5% of the markers.

In Q1, 34 RGAs, generating 73 markers were mapped. The population has been phenotyped over two years for common rust resistance and Pachymetra resistance. Both parents have an intermediate level of resistance to the two diseases. In the male map (74C42), one RGA was identified as associated with rust resistance over both years with two others associated in one year only, and are therefore questionable. In the female map, 2 RGAs were shown to be associated with rust resistance over both years, with an additional 2 RGAs associated in one year only. For Pachymetra resistance, no RGAs were shown to be associated over both years data. Thus 3 of the 73 RGA markers (4%) were associated with rust resistance. This compares favourably with 3% of the AFLP and SSR markers.

4. **Use the "candidate genes" as anchor points to compare maps and correlate linkage groups in the 2 populations.**

Partly Achieved:

Due to technical difficulties and time limitations, the candidate genes associated with sugar- and fibre- related traits were not mapped in Q1 or Q3. Thus, we have not been able to use "candidate genes" as anchor points to compare maps. Instead, we have used microsatellite markers (SSRs) as our anchor points to enable map and QTL comparisons between the Australian and French sugarcane populations.

Approximately 60 SSRs markers have been mapped in R570 of which 31 have also been mapped in Q3. Markers generated by forty-eight microsatellites have been shown to be associated with fibre and sugar-related

traits in R570. Seventeen of these SSRs are also associated with these traits in Q3, with 9 SSRs associated from each parent. The association of these 17 SSRs with similar traits in both R570 and Q3 populations provides extra confidence in the reliability of the association and the usefulness of the SSR marker. It also highlights the usefulness of comparative mapping for populations with smaller maps, such as Q3, because it identifies additional genomic regions associated with the target traits. The SSRs associated with these traits in R570 but not mapped in Q3 should be the focus of future marker tagging activities.

5. **Investigate methods for more rapidly screening EST markers in mapping populations.**

Achieved:

ESTs can be successfully used as RFLP probes and we have used this approach to map many ESTs and RGAs onto sorghum and sugarcane. However, RFLPs in sugarcane are tedious, time consuming and technically difficult, especially considering the thousands of ESTs available for use. We have investigated three approaches for their ability to increase the efficiency of utilisation of ESTs. The three techniques included:

a) SSRs within ESTs. Only approximately 3% of CSIRO/SRDC ESTs contained SSR sequences. Primers were designed to 35 EST-derived SSRs and tested on sugarcane varieties and ancestral species. Forty percent of the SSRs were polymorphic between sugarcane varieties, but this level is approximately half that achieved by genomic SSRs. An advantage of these EST-derived SSRs over genomic SSRs was that a higher percentage of the former were usable and polymorphic in related genera such as *Erianthus* sp.

b) primers designed to sequence variation between homologues. Two sets of primers were designed to 8 ESTs and tested on the parents of the Q1 population and 5 progeny. Both sets of primers were monomorphic and non-segregating for 3 of the 8 ESTs. At least one set of primers was polymorphic for the remaining 5 ESTs, but primers for only 1 EST amplified a product that was obviously segregating in the progeny.

c) real-time PCR using primers designed to sequence variation between homologues. One primer pair designed to one of the above 8 ESTs was tested on 5 genotypes. The result demonstrated the usefulness of this approach to detect dosage variation in genotypes and thus may be useful when mapping important ESTs where the variation is not single-dose.

Three different techniques have all have proved successful. However, none of the techniques enable large numbers of ESTs to be mapped. They are more useful for mapping a subset of ESTs.

METHODOLOGY

- **Phenotypic Trait Evaluation**

Three sugarcane populations have been evaluated in this project, namely the French population R570 and two Australian populations, Q1 (Q117 x 74C42) and Q3 (Q117 x MQ77-340). Phenotypic information for R570 (Brix, height, stalk diameter and sucker number) was generated by Dr Jean-Yves Hoarau, CERF, Reunion Island. Phenotypic information for Q1 (rust resistance and Pachymetra resistance) was obtained as part of CTA049.

Phenoytpic information for Q3 (Brix, Pol, CCS, and fibre) was obtained by Dr Phil Jackson, CSIRO Plant Industry, Townsville. The Q3 population has 232 progeny. The progeny and parents were planted at two sites in northern Australia using local industry practices. The trial design was a randomised complete block design with 4 replicates. Individual plots were 1 row x 5 m and the row spacing was 1.5 m. Sampling for fibre, CCS, Pol, Brix and purity was done each year, using four samples from each plot. Measurements followed standard Australian sugar industry procedures (BSES, 1984).

- **Genotypic information**

DNA of the Q3 progeny was extracted using the CTA method described by Hoisington, 1992.

RGAs: All techniques relevant to RFLP analysis; including restriction enzyme digestion, probe preparation, hybridisation and autoradiography were carried out as described previously in BSCS1 (Final Report submitted to SRDC) and the Laboratory Protocols Manual developed by CIMMYT (Hoisington, DA. Laboratory Protocols. CIMMYT Applied Molecular Genetics Laboratory. Mexico, D.F.).

SSRs: All techniques relevant to SSR analysis were carried out as described in Cordeiro et al, 2000.

AFLPs: All techniques relevant to AFLP analysis were carried out as described in Vos et al, 1995.

For Q3, maps were constructed using Mapmaker 3.0 (Lander et al, 1987). The male and female maps were constructed initially with the 1:1 markers, with the 3:1 markers added later. Two point analyses were performed at a LOD of 5 and a recombination fraction of 0.35. Single factor QTL analysis was conducted using Map Manager QTXb16 (Meer et al, 2002).

For the incorporation of RGA markers into the Q1 map, Map Manager was used for both the initial linkage analysis and single-factor analysis.

For sorghum, RFLP analysis and map construction were carried as described in Tao et al, 1998.

RESULTS

1. **To identify and request "candidate genes" in the sucrose and carbohydrate metabolism, fibre synthesis and disease resistance pathways. Search EST databases, especially the SRDC/CSIRO sugarcane EST database, for "candidate" genes.**

Achieved:

The CSIRO/SRDC EST database contains approximately 8500 ESTs. The sequence of each EST has been checked for homology to other genes in public domain databases and a tentative identity, based on sequence homology, obtained for approximately 55% of them. Approximately 20% of the ESTs are homologous to other genes with no known function and the remaining 25% are not significantly homologous to anything currently isolated. Of the approximately 4500 ESTs with a tentative identity (55% of 8500 ESTs), approximately 400 were identified as potentially relevant. These ESTs were classified as members of sugar and fibre biosynthetic pathways (sugar metabolism and fibre-related protein classes), homologous to known resistance genes or resistance mechanism genes (stress-related class), or known regulators of gene expression (gene expression, signal transduction and membrane classes). The approximately 400 ESTs (Table 1) were screened over the parents of sugarcane and sorghum mapping populations. In addition, 55 resistance gene analogues (isolated in a CRC Tropical Plant Protection project) were also identified.

Table 1: Number of candidate genes screened in sorghum and sugarcane mapping populations.

EST Class	**No. clones screened**
Sugar metabolism	42
Fibre-related protein	48
Gene expression	60
Signal transduction	52
Membrane	60
Stress-related	97
Other	43
Total EST	**396**
Resistance Gene Analogues	55

Polymorphism levels varied between 39 and 60% for the parents of the 3 sorghum mapping populations. Currently, approximately 100, 150, and 150 sugarcane ESTs have been mapped in the 3 sorghum populations. These ESTs cover the range of categories listed above. The accompanying figure (Figure 1) shows the location of approximately 150 sugarcane ESTs in one of our sorghum mapping populations and the relative position of QTLs for resistance to 3 sorghum diseases. Of particular interest is the observation that a sugarcane disease resistance candidate gene with homology to a

major maize rust resistance gene maps to a major QTL for rust resistance in sorghum.

In sugarcane, polymorphism level between the parents of the BSES sugarcane map, Q1, was approximately 49% and similar polymorphism levels have been seen in Q3. Due to technical difficulties and time constrains, these candidate genes have not been mapped further in Q3.

The 55 RGAs were screened over parents of the sorghum and sugarcane mapping populations. In sorghum, 31 of the 55 RGAs were polymorphic in the most polymorphic sorghum mapping population and all 31 have been mapped. In sugarcane, 34 of the 55 RGAs gave single-dose polymorphisms. The 31 clones yielded 77 markers and have been included in the Q1 map. Given the complexity of the sugarcane genome, the sorghum map (2x=2n=20 chromosomes) is providing a useful reference point for simplifying our sugarcane maps.

2. **Map "candidate genes" in 2 populations (one from BSES and the other from CIRAD) that are segregating for CCS, fibre and disease resistance.**

Achieved:

Candidate genes for sugar and fibre:
Thirty-two putative candidate genes (Table 1) were screened across 112 progeny of R570 of which twenty-three gave scorable single-dose patterns. These 23 genes generated 139 markers of which 83 were single dose. Sixty-five of the 83 markers have been included in 35 co-segregation groups in the map of R570; the remaining 18 markers are unlinked. These candidate genes include genes from the carbohydrate and fibre biosynthetic pathways.

Although the above 32 candidate genes were also screened across parental enzyme blots of Q3 and several attempts were made to screen progeny of the Q3 population, no putative candidate genes have been mapped to date in Q3, due to technical and time limitations experienced by Ms Nathalie Reffay during her stay in CSIRO and her recent visit to CIRAD.

Candidate genes for disease resistance:
Rust resistance in R570 has previously been reported to be due to a major gene and several RFLP probes were shown to be linked to this gene, including CDSR29, SSCIR194 and TXS1257, with CDSR29 being the closest marker. These 3 RFLP probes were screened across progeny of the Q1 (Q117 x 74C42) mapping population. Both CDSR and SSCIR194 gave scorable single-dose markers (4 and 3, respectively). One hundred progeny of Q1 have been scored for common rust over 2 years as part of CTA049.

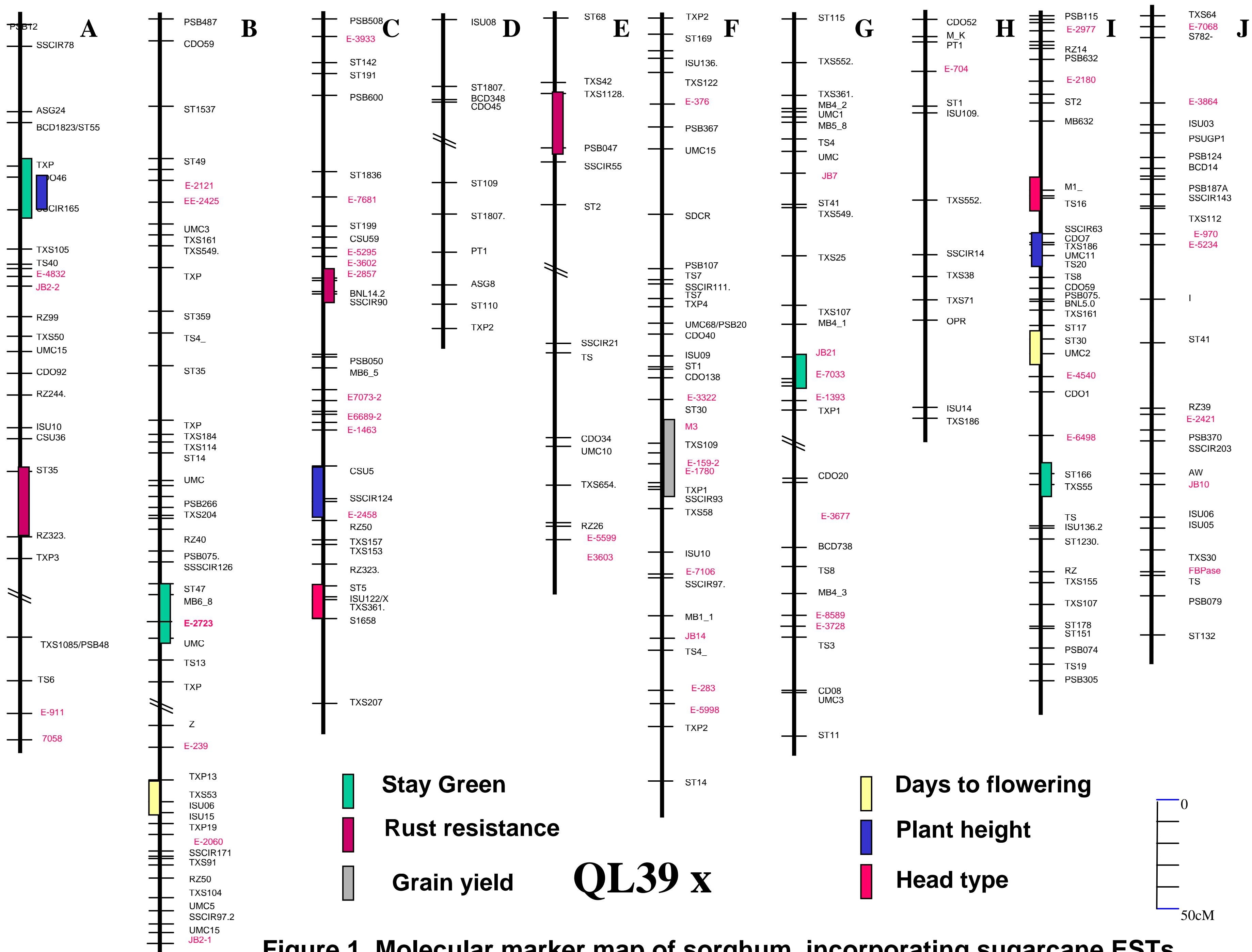


Figure 1. Molecular marker map of sorghum, incorporating sugarcane ESTs.

Table 2. List of candidate genes tested for polymorphism in R570 and their origin

	Fibre- associated Enzymes	Mapped	Source
1	cellulose synthase		CSIRO
2	cinnamate 4-hydroxylase		CSIRO
3	cinnamyl alcohol dehydrogenase		CSIRO
4	laccase		CSIRO
5	caffeoyl-CoA 3-O-methyltransferase		CSIRO
6	Peroxidase		CSIRO
	Sugar-associated Enzymes		
1	cell wall invertase		CSIRO
2	Fructose-6-phosphate 2-kinase/fructose-2,6-bisphophatase	no	CSIRO
3	fructose-bisphotate aldolase		CSIRO
4	sucrose phosphate synthase	no	CSIRO
5	glucose -induced repressor protein	no	CSIRO
6	phosphofructo-1-kinase		CSIRO
7	sucrose-induced protein	no	CSIRO
8	UDP-glucose dehydrogenase		CSIRO
9	4-alpha-glucanotransferase		CSIRO
10	6-phosphogluconate dehydrogenase	no	CSIRO
11	glucose-6-phophate dehydrogenase	no	CSIRO
12	glucose-6-phosphate isomerase	no	CSIRO
13	glyceraldehyde-3-phophate dehydrogenase		CSIRO
14	ribulose-phosphate 3-epimerase		CSIRO
15	trehalose-6-phosphate synthase	no	CSIRO
16	Enolase	no	Sth Africa
17	Sucrose synthase		Sth Africa
18	Triose P isomerase		Sth Africa
19	Sucrose synthase		Sth Africa
20	Vacuolar invertase		France
21	3'ADPGpyrophosphorylase, -shrunken2		France
22	5' ADPGpyrophosphorylase, -shrunken2		France
23	ADPGpyrophosphorylase-Brittle2		France
24	Debranching snzyme, Sugary1		France
25	B-galactosidease		France
26	Opaque2		France

3. **Measure correlation between CCS, sugar yield, fibre and disease resistance and "candidate genes" in order to see if the candidate genes provide "perfect" or better markers for these traits than current marker techniques.**

Achieved:

In all populations tested, candidate genes consistently gave a higher probability of finding an association with a trait, than did anonymous RFLP, SSR or AFLP markers.

Candidate genes in R570

In R570, the 65 markers generated by the 23 candidate genes (Table 2) were screened across 112 progeny from the population. The information was analysed using single-factor analysis and results obtained for the 4 traits (Brix, height, stalk number and sucker number). The phenotypic information was obtained by Dr Jean-Yves Hoarau on Reunion Island for these 4 traits for all 300 R570 progeny, including the 112 assessed here. Of the 4 traits assessed for all 112 progeny, only 4 markers were consistently associated with a trait over the two years for which data was collected (Table 3). One marker was identified as associated with Brix (BRX95 and 96), another was identified as associated with height (HT95 and 96) and two were identified as associated with stalk diameter (DIA95 and 96). No markers were consistently identified with sucker number. These significant markers represent 6% of the candidate gene markers tested. By contrast, Hoarau et al (2002) found 40 markers from a total of 1000 AFLP markers to be associated with these 4 traits, of which only 5 appeared in both years. This represents only 0.5% of the markers.

Table 3 – Candidate genes associated with Brix (BRX), stalk diameter (DIA), height (HT) and sucker number (SUCK) measured for two years in R570

Trait	Phospho-fructo-1-kinase	Ribulose-phosphate 3-epimerase	UDP-glucose dehydrogenase	ADPG pyro-phosphorylase	Cinnamyl alcohol de-hydrogenase	Cell Wall Invertase
BRX95		***	**			
BRX96			*			
DIA95				**		*
DIA96				*		****
HT95	*				*	
HT96					***	
SUCK95						
SUCK96						

* P=0.05, ** P=0.01

Candidate genes in Q1

In Q1, 34 RGAs, generating 73 markers were mapped. The population has been phenotyped over two years for common rust resistance and Pachymetra resistance. Both parents have an intermediate level of resistance to the two diseases.

In the male map (74C42), 3 different RGAs were associated with rust resistance but only one (DR-T2) was associated with resistance over both years (Table 4). The two other RGAs (including one with strong homology to a major maize rust resistance gene – DR-8167) were only associated with rust resistance in one year, and are therefore questionable. In the female map, six different markers generated by 4 different RGAs were shown to be associated with rust resistance. Again, only 2 markers were shown to be associated with rust resistance over both years. One of the markers was generated by a RGA

which was also associated with resistance in the male map in 2000 (DR-22.1.5).

For Pachymetra resistance, four markers generated by 3 RGA probes in the male map (74C42) were shown to be associated with Pachymetra resistance using data from the first trial, but none were associated with the second year's data, suggesting that the association in year 1 was a statistical artefact or environment-specific (Table 4). No RGA markers from the female map were found to be associated with Pachymetra resistance. Thus 3 of the 73 RGA markers (4%) were associated with rust resistance. This compares favourably with 14 of 445 (3%) markers in the male map and 12 of 407 (3%) markers in the male map associated with rust resistance for AFLP and SSR markers.

Thus, although the number of candidate gene markers used is small, the results suggest that candidate genes are more likely to identify markers linked to the targeted traits, than random markers such as AFLPs.

Table 4 – Resistance Gene Analogues associated with Disease Resistance in Q1

Parent	**Rust (00)**	**Rust (01)**	**Pachy. (01)**	**Pachy. (02)**
Female	DR-43.1.5a**			
(Q117)		DR-22.1.5a*		
		DR-22.1.5b*		
	DR-22.1.5c*	DR-22.1.5c*		
	DR-7R1.3-2*			
	DR-40.1.16a*	DR-40.1.16a*		
Male			DR-40.1.8b**	
(74C42)			DR-18.25H**	
			DR-29.3.1-1a*	
			DR-29.3.1.1c*	
	DR-8167-1*			
	DR-22.1.5a*			
	DR-T2-2*	DR-T2-2*		

4. **Use the "candidate genes" as anchor points to compare maps and correlate linkage groups in the 2 populations.**

Partly Achieved:

Comparative mapping of regions associated with sugar-related traits
Thirty-two putative candidate genes were screened across 112 progeny of R570 of which 23 gave scorable single-dose patterns. These 23 genes generated 139 markers of which 83 were single dose. Sixty-five of the 83 markers have been included in 35 co-segregation groups in the map of R570; the remaining 18 markers are unlinked. These candidate genes include genes from the carbohydrate and fibre biosynthetic pathways (Table 2).

Although the above 32 candidate genes were also screened across parental enzyme blots of Q3 and several attempts were made to screen progeny of the

Q3 population, no putative candidate genes have been mapped to date in Q3, due to technical and time limitations experienced by Ms Nathalie Reffay during her stay in CSIRO and her recent visit to CIRAD. Thus, we have not been able to use "candidate genes" as anchor points to compare maps. Instead, we have used microsatellite markers (SSRs).

Approximately 60 SSRs from the International Consortium of Sugarcane Biotechnology have been mapped in R570 of which 31 have also been mapped in Q3. In R570, markers generated by forty-eight microsatellites have been shown to be associated with fibre and sugar-related traits. Seventeen of these SSRs are also associated with these traits in Q3, with 9 SSRs associated from each parent (Table 5).

The shaded boxes illustrate where the same trait is associated with the SSR in R570 and one of the Q3 parental maps (Q117 is the female parent and MQ77-340 is the male parent). Surprisingly, allelic markers from only 1 SSR were associated with one of these traits in both parents of the Q3 population (markers from SSR ms15 for fibre) which may be due to the relatively small number of markers mapped in each parent.

The association of these 17 SSRs with similar traits in both R570 and Q3 populations provides extra confidence in the reliability of the association and the usefulness of the SSR marker. It also highlights the usefulness of comparative mapping for populations with smaller maps, such as Q3, because it identifies additional genomic regions associated with the target traits. The SSRs associated with these traits in R570 but not mapped in Q3 should be the focus of future marker tagging activities.

<u>Comparative mapping of regions associated with rust resistance</u>
Rust resistance in R570 has previously been reported to be due to a major gene and several RFLP probes were shown to be linked to this gene, including CDSR29, SSCIR194 and TXS1257, with CDSR29 being the closest marker. These 3 RFLP probes were screened across progeny of the Q1 (Q117 x 74C42) mapping population. Both CDSR and SSCIR194 gave scorable single-dose markers (4 and 3, respectively). One hundred progeny of Q1 have been scored for common rust over 2 years as part of CTA049. As presented in CTA049, the distribution of rust resistance in the Q1 population differs markedly from the distribution in R570, suggesting that rust resistance in the Q1 population is quantitative and not controlled by a single gene.

Single factor analysis was undertaken using the single-dose RFLP markers. None of the 4 single-dose RFLP markers generated by the CDSR29 probe were associated with rust resistance. CDSR29 markers are linked to AFLP markers in two linkage groups and the AFLP markers were also not associated with rust resistance. One of the single-dose RFLP markers generated by probe SSCIR194 was shown to be weakly associated with rust resistance for one year's data only. As the marker was not significantly associated with rust resistance in the second year, the association of this marker with rust resistance is questionable.

Table 5 – Microsatellite markers associated with sugar and fibre- related traits in R570 and Q3 populations.

		ms8	ms10	ms12	ms14	ms15	ms17	ms18	ms27	ms28	ms34	ms35	ms36	ms39	ms42	ms52	ms53	ms54
R570	**BRX**	X		O		X	O	X	O		X	X	O	O	O	X	O	X
	HT			O	O	O			X			X						O
	DIA			O	X				O	O			O					O
	SUCK				X				O									O
MQ77-340	**CCS**		X						O			O	O					
	PUR		X						O									O
	FBR					O	O	O				O		O				
	POL		X															
	BRX		X			O												
Q117	**CCS**	X		O						O	X				O		O	
	PUR	O								O					O			
	FBR	O			O	O												
	POL			O						O							O	
	BRX			O											O	O		

X = > 0.01
0.5 > O > 0.01

The lack of association between CDSR29 and SSCIR194 markers and rust resistance in the Q1 population could be due to several reasons. Firstly, it may be that rust resistance in Q1 is due to a different rust resistance gene or genes to the resistance gene in R570. Secondly, only 4 and 3 single dose markers were generated by CDSR29 and SSCIR194 probes, respectively, in our population. Thus our RFLP analysis only allowed us to examine the association between a small number of alleles of a possible 10-12 alleles generated by the two RFLP probes. Rust resistance in R570 is a single gene. It is possible that the CDSR29 and SSCIR194 alleles linked to the single gene in R570 were not polymorphic or single-dose in the Q1 population and thus we could not identify the association between these alleles and rust resistance. However, the distribution of rust resistance in the Q1 population is more consistent with a quantitative trait, as discussed previously in CTA049, and not a single gene, suggesting that the single gene of R570 is not present in this population or not effective against the rust strain present in Australia.

<u>Comparative mapping of RGAs in sugarcane and sorghum</u>

Thirty-four RGAs have been mapped in the Q1 sugarcane population and 23 have been mapped in sorghum, of which 13 are common to both maps. Of the 34 RGAs mapped in both sugarcane parental maps, 3 linkage groups from the female map and 3 linkage groups from the male map contained 2 or more RGAs. For one female linkage group and one male linkage group, two linked RGAs have also been shown to be linked on the same linkage group in sorghum (Figure 2); these two linkage groups are the only sugarcane linkage groups where two RGAs have been mapped in sorghum and both illustrate the syntenic location of these two pairs of RGAs in sorghum and sugarcane. This again demonstrates the usefulness of sorghum as a genetic reference for its more complex relative.

One of the RGAs mapped in sorghum has strong homology to the Rp1 maize rust resistance gene. In sorghum, this RGA (DR-8167) maps to a major rust resistance QTL. This RGA was also mapped in sugarcane. It was found to be weakly associated with rust resistance in one year only. Curiously, in the sugarcane male map, the only RGA which was consistently associated over the two years of data, DR-T2, maps to the same sorghum linkage group as DR-8167, but at the opposite end.

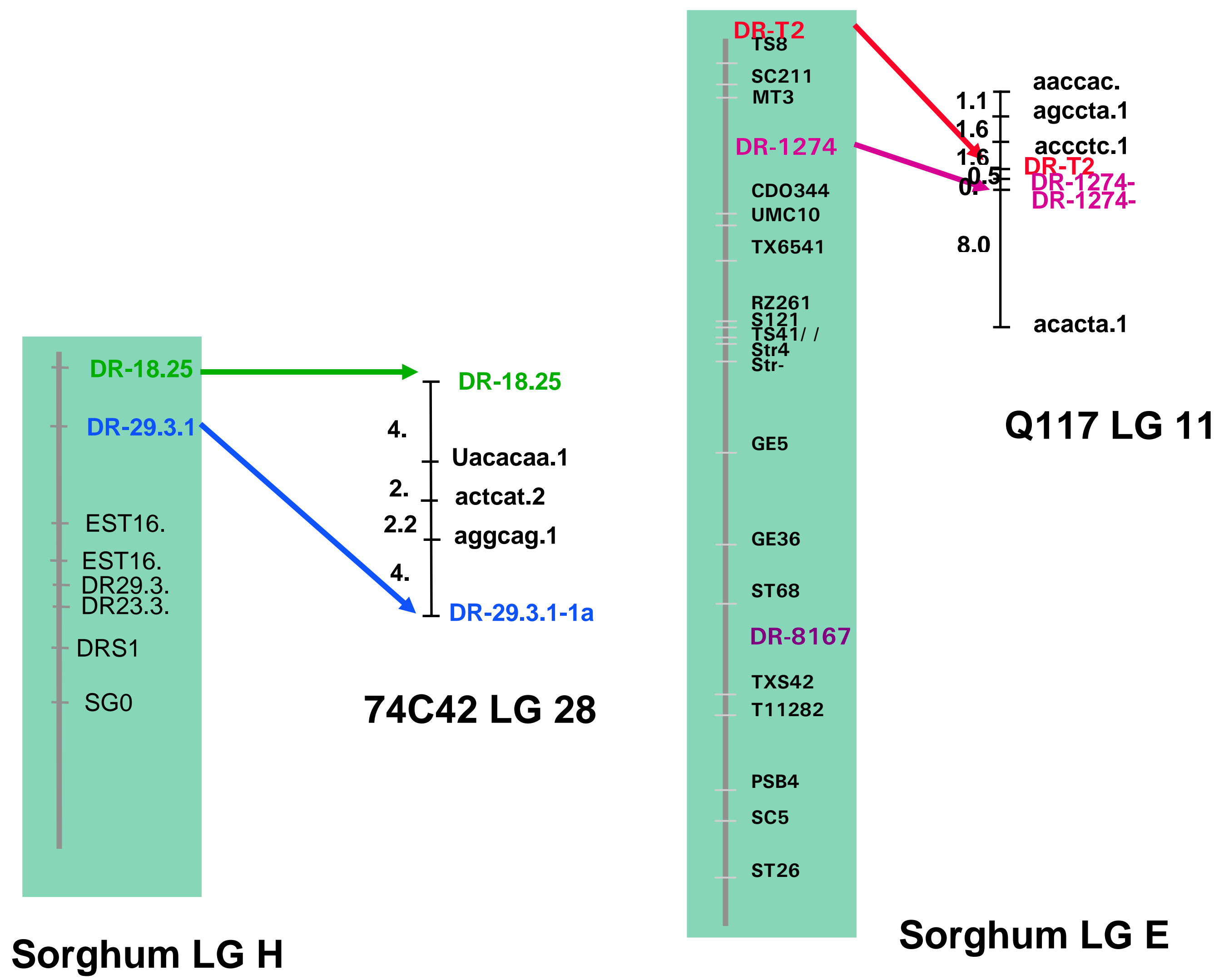
CIRAD have recently isolated and mapped approximately 75 RGAs from sugarcane. As part of my visit to CIRAD later this year, it will be interesting to compare RGAs isolated and their map position. It should enable linkage groups in Q1 and R570 to be aligned.

5. **Investigate methods for more rapidly screening EST markers in mapping populations.**

Achieved:

ESTs can be successfully used as RFLP probes and we have used this approach to map more than 100 ESTs onto sorghum (see Milestone 3).

Figure 2 – Syntenic position of RGAs (DR29-3-1-1 and DR18.25, and DR-T2 and DR-1274) in sorghum and sugarcane (Q1) populations



However, this is tedious and time consuming, especially considering the thousands of ESTs available for use. We have investigated several approaches to increase the efficiency of utilisation of ESTs. These methods and their advantages and disadvantages are discussed.

1. Identification of SSRs within ESTs

In collaboration with Dr Giovanni Cordeiro at SCU, the CSIRO EST collection was screened for microsatellite sequences. More than 250 microsatellite sequences were identified in the 8,678 ESTs. A subset of 35 SSR-containing ESTs were identified. These ESTs contained SSRs of adequate length (>5 repeats) and had sufficient surrounding sequence information to enable primers to be designed.

In contrast to microsatellites derived from sugarcane genomic DNA, the EST SSRs were shorter and mainly trinucleotides and were always perfect repeats.

Primers were designed to the 35 SSRs and tested on five genotypes (Q124, Q117, R570, *S. officinarum* Chittan and *S. spontaneum* Saigon). Of the 35, 21 produced a PCR product and 17 were polymorphic in 2 or more of the 5 genotypes tested. Eight were polymorphic within the 3 sugarcane genotypes and 16 were polymorphic between at least one variety and an ancestral *Saccharum* species. However, only a small number of alleles were detected, with a maximum of 5.

Five of the 17 primers were also tested on 4 *Erianthus* species and 3 *Sorghum* species. All five cross-transferred and were polymorphic. This is in contrast to genomic SSRs which rarely transfer to other species and show very low levels of polymorphism.

This study suggests that it would be possible to map and utlise ESTs containing SSR sequences by designing primers to flank the ESTs. While polymorphism levels are lower within sugarcane using these EST SSRs rather than genomic SSRs, it would be possible to map a subset. In addition, these SSRs would be useful to provide map anchor points between sugarcane and related species.

2. Identification of length polymorphisms using primers targetted to regions of sequence variation in EST multigene families.

Examination of the EST database revealed multiple copies of some sequences. When the sequences were aligned, it could be seen that many of the copies were identical, suggesting the same gene, while others contained sequence variation and were allelic variants or members of multigene families. The sequence variation included deletions/insertions and point mutations.

Two sets of primers were designed to each of 8 ESTs, one pair to detect length polymorphisms and one pair to detect specific alleles (Table 6). The primers were tested on the parents of the Q1 mapping population (Q117 x 74C42).

Table 6 –Sugarcane ESTs to which primers were designed to identify length polymorphisms for mapping

EST	Polymorphism Result
Transcription factor 2	both primer pairs were monomorphic
Transcription factor 4	one primer pair was polymorphic and segregated 1:1 in the progeny as expected. The other primer pair was monomorphic
Transcription factor 5	both primer pairs were polymorphic
Transcription factor 6	both primer pairs were monomorphic in the parents
Remorin	both primers were polymorphic, but missing lanes made the data inconclusive. Non-specific bands are amplifying
Xyloglucan endotransglycosylase	one primer pair appeared to amplify in one parent only. The other primer pair was polymorphic
Serine/threonine kinase	both primer pairs were monomorphic
No Known Function	both primer pairs were polymorphic

None of the primers designed to detect specific alleles appear to do so. In all cases, an amplification product was seen in both parents. These amplification products may result from alleles at other chromosomes in this highly polymorphic species, suggesting that it is difficult to design these primers if all the alleles of a locus are not known. Another possibility is that the specific primer may need more mismatches to increase its specificity. This will be tested later.

Of the primers designed to detect length polymorphisms, five of the 8 primers detected polymorphisms. One of these primers was screened across 18 progeny and segregated 1:1, as expected. This EST will be mapped onto our sugarcane map.

This approach appears to give limited success and may be due to the ESTs being from a single genotype. Thus the sequence variants in multigene families targetted in this study may be conserved in other sugarcane genotypes. More polymorphic primer combinations will be tested in progeny of the Q117 x 74C42 cross. This approach will also be re-evaluated using homologous ESTs from different genotypes, as this information becomes available.

3. Use of Real-time PCR assays.

The Taqman approach was tested using one of the above primer sets which only amplified a single product. The objective of this approach was to see if this technique could detect allele dosage differences in 5 different genotypes (Q117, H56-752, Q138, Coimbatore and Badila). Three different DNA extractions were used for each cultivar.

The results suggested that the genotypes had different numbers of alleles detected by the primers, however, there was considerable variation in the different genotype extracts. Thus, it would appear that real-time PCR assays can detect large dosage differences. This technique will be further investigated and modified to improve technical reproducibility.

4. Conclusions and future directions.

Three different techniques have been attempted and all have proved successful. However, none are suitable for rapid screening of large numbers of ESTs. These techniques would be suitable for mapping of specific and important ESTs only. Further work is ongoing to identify new techniques to improve the efficiency of mapping with candidate genes.

PROJECT OUTPUTS

The project has demonstrated that:

- Candidate genes are a useful source of markers, providing information on the potential contribution of individual genomic regions to expression of traits of interest.
- Candidate genes have been identified that are associated with disease resistance, and sugar- and fibre- related traits.
- There are numerous techniques for mapping ESTs, all of which have advantages and disadvantages. However, RFLPs remain the best technique for screening large numbers of ESTs. Other techniques showed promise for identifying larger numbers of alleles of individual genes/ESTs of interest
- Maps and QTLs can be aligned using candidate genes and SSR markers.
- Sugarcane polyploidy limits the usefulness of candidate genes as it is difficult to test every sugarcane gene allele/locus for linkage to a trait.
- Sugarcane polyploidy and chromosome number makes alignment of maps and QTLs difficult as each map and hence QTL is likely to be targeting overlapping, but not identical, regions of the sugarcane genome.

PROJECT OUTCOMES

This project was established to investigate the usefulness of candidate genes as markers in sugarcane and to identify techniques for efficient utilisation of ESTs as candidate genes. The project has generated useful insights that will influence the use of ESTs as candidate genes in future mapping projects.

Firstly, although only a small number of candidate genes have been tested, it would appear that candidate genes are more efficient at generating markers than are random markers such as AFLPs and SSRs. A higher percentage of linked markers was obtained using candidate genes in both R570 and Q3, than was obtained using random markers.

Secondly, none of the current techniques appear to be suitable for screening of large numbers of ESTs as candidate genes. The complexity of the sugarcane genome appears to be a major limitation in mapping markers of any kind as only markers segregating 1:1 can be evaluated. In addition, sugarcane contains between 9-13 doses of every chromosome, and thus only a limited number of chromosomes containing the target genes can be evaluated in any one cross. RFLPs, in our hands, usually only detect 2-3 homologous chromosome, whereas SSRs usually detect at least 4. Thus, SSRs are more effective at tagging homology group chromosome members. They are also much easier to detect than RFLPs.

Thus, candidate genes are a useful source of markers. However, current methods to evaluate individual candidate genes, while effective, are still laborious.

FUTURE RESEARCH NEEDS

Several research needs have been identified in this project.

- Several candidate genes have been identified that are associated with disease resistance, and sugar- and fibre- related traits. Further research will be required to determine whether the association is coincidental, in which case the candidate gene is still a useful marker, or whether the gene is implicated in the expression of the trait, in which case it could provide insight into the underlying genetics and biochemistry/physiology of the trait itself.
- Of the techniques investigated, RFLPs enable the largest number of ESTs to be screened for use as candidate genes. However, given the technical difficulties associated with the use of RFLPs in sugarcane, more work is required to identify and test other techniques.
- SSRs are very useful markers for map comparisons as they are very polymorphic, enabling more chromosome members of a homology group to be tagged, and easy to use. Current maps only contain a subset of available sugarcane markers. A reference map containing all of the ICSB markers would facilitate sugarcane map comparisons. There are also a large number of sorghum and maize SSRs. Our lab has shown that many of them are mappable in sugarcane. A sugarcane reference map containing sorghum and maize SSRs would also facilitate comparative mapping with these simpler and more extensively studied relatives.

RECOMMENDATIONS

Numerous recommendations arise from this project.

1. Genome coverage is important when comparing chromosome regions. Thus PCR markers which are easier and faster to generate than RFLPs are advisable. Despite large marker numbers, most sugarcane maps have less than 50% genome coverage and therefore cover different regions of the genome. This makes it very difficult to compare genomic regions associated with specific traits as different chromosomes members of a homology group are likely to have been mapped in different maps. Most genome comparisons can only be made at the homology group level. Nevertheless, this information is useful and all maps should contain SSRs to enable some comparisons to be made.

2. RFLPs are technically demanding and tag fewer members of homology groups than do SSR markers. Thus, SSR markers are more useful in map comparisons than RFLP markers. However, while SSRs can be compared between Andropoganae species (maize and sorghum), RFLP markers will allow map comparisons across a much wider range of species, including all members of the grass family.

3. Candidate genes are useful markers, however, current techniques do not allow rapid screening of large numbers of markers. Current techniques are effective at mapping smaller subsets of genes, or in mapping allelic variants of specific genes. New methods of evaluating ESTs and candidate genes should be attempted as they are developed.

PUBLICATIONS arising from the project

Cordeiro, GM, Casu, RE, McIntyre, CL, Manners, JM and Henry, RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs. Plant Science 160: 115-1123.

McIntyre, C.L., Drenth, J., Knight, D., Jordan, D., Casu, R. and Manners, J. Comparative mapping of resistance gene analogues in sorghum and sugarcane
In: Plant Breeding for the 11th Millenium, J.A.McComb (ed), Proceedings 12th Australasian Plant Breeding Conference (Perth, 15-20 September, 2002) p 96-99

Reffay, N, D'Hont, A, Besse, P. and McIntyre, C.L. Comparative mapping between Australian and French mapping populations identified both new and common sugar-related QTLs. In prep.

Publications cited in the Final Report

Cordeiro G.M., Taylor G.O., Henry R.J., 2000. Characterisation of microsatellite markers from sugarcane (*Saccharum* sp.), a highly poluploid species. Plant Science 155: 161-168.

Hoarau JY, Grivet L, Offmann B, Raboin L-M, Diorflar J-P, Payet J, Hellmann M, D'Hont A, Glaszmann JC, (2002) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.).II. Detection of QTLs for yield components. Theor Appl Genet.

Hosington D.A., 1992. Laboratory protocols. CIMMYT Apllied Molecular Genetics Laboratory. Mexico, D.F. CIMMYT.

Lander E.S., Green P., Abrahamson J., Daly M.J., Lincoln S.E., and Newburg L. 1987. MAPMAKER: an interactive computer package for construction primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

Meer, JM, Manly, KF, Cudmore, RH, 2002. Software for genetic mapping of mendelian markers and quantitative traits loci. Roswell Cancer Park Institute.

Tao, YZ, Jordan, DR, Henzell, RG and McIntyre, CL 1998. Identification of genomic regions for rust resistance in sorghum. Euphytica 103: 287-292.

Vos P., Hogers R., Bleeker M., Reijans M., van der Lee T., Hornes M., Fritjers A., Pot J., Peleman J., Kuiper M. And Zabeau M., 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acid Res., 23: 4407-4414.