FINAL REPORT
PROJECT ICB010

Project Title	***Validation of Single Nucleotide Polymorphisms (SNPs) in Sugarcane ESTs as Useful Genetic Markers***
Project Number	ICB010
Principal Investigators	*Dr Giovanni Cordeiro & Prof Robert Henry* Centre for Plant Conservation Genetics Southern Cross University PO Box 157 Lismore, NSW 2480 Australia *Phone*: 02-6620 3356 *Fax*: 02-6622 2080 *E-mail*: gcordeir@scu.edu.au rhenry@scu.edu.au
Milestone No	4 – Final report on evaluation of SNP markers in sugarcane.
Milestone Target Date	31 June 2004
Milestone Achievement Date	20 January 2004

This project has been funded by the Sugar Research and Development Cooperation to the amount of **AUD**21,350.00**.**

Confidentiality:

SRDC and SCU acknowledge that, for a period of five years after the completion of this project, it will keep confidential any information disclosed, including Completed SNP Primer sequences obtained, other than information in the lawful possession and control of the SRDC prior to the start of this project

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 this Project.

Executive Summary

Sugarcane is one of the most important field crops grown in the tropics and sub-tropics. It possesses a highly complex genome, with multiple copies of a single gene present. Breeding of improved cultivars of sugarcane is difficult because most traits are derived from multiple genes and are quantitatively inherited. The complexity of the sugarcane genome makes sugarcane breeding and genetic analysis a challenge. Commercial sugarcane plants are the result of a limited series of crosses and backcrosses derived from the domesticated species *Saccharum officinarum* L. (2n = 80) and the wild species *S. spontaneum* (2n = 40 – 120). As a result of this process, commercial sugarcane plants are inter-specific poly-aneuploid hybrids with chromosome numbers usually in excess of 100.

Single nucleotide polymorphisms (SNPs) are an alteration of one nucleotide in a DNA sequence. These single changes can be detected and used as markers, and because their occurrence is frequent, they provide a large source of genetic markers. SNP markers represent the smallest possible genetic difference and the largest source of genetic markers. This project takes advantage of these nucleotide differences as a new strategy for analysis of traits in sugarcane, and also takes advantage of new technologies to measure these sequence variations. The marker system makes use of the Pyrosequencer, a sophisticated instrument that uses, through a series of biochemical reactions, light emissions to detect these differences in the genetic sequence and ultimately determines the different proportions of alleles present for a particular gene. The application of this technology although not new to organisms carrying only a diploid genome such as humans, has never been applied to a genetic system with such a high level of complexity as sugarcane.

Through a collaborative effort with the International Consortium for Sugarcane Biotechnology (ICSB), an initial set of 33 markers have been developed. We have now shown that with the aid of these markers, it is possible to utilise these minor differences in gene sequence to deduce the allele content in different sugarcane genotypes. The technology is able to identify the origins of an allele in commercial canes, indicating its history from either a *Saccharum officinarum* or *S. spontaneum* parent. It also provides a 'fingerprint' of different sugarcane genotypes that allows the breeder to determine the genetic relatedness of any two varieties. However, the most powerful application of the SNP marker system will be in the identification of combinations of SNPs in a gene sequence which act to identify individual gene haplotypes that may be considered as the allele equivalents in the sugarcane genome. This would provide sugarcane breeders with the means to select parents with the greatest number of desirable gene haplotypes to maximise the number of offspring carrying the desired trait. Hence, it is now possible to identify candidate gene sequences that may underpin important traits in sugarcane and to characterise single nucleotide polymorphisms (SNPs) in these genes.

BACKGROUND

Single nucleotide polymorphisms (SNPs) are an alteration of one nucleotide in a DNA sequence. These single changes can be detected and used as markers, and because their occurrence is frequent, they provide a large source of genetic markers. SNP markers represent the smallest possible genetic difference and the largest source of genetic markers. In addition, they are more likely to be located close to or within the target gene. SNP markers have become extremely valuable in human genetics and are beginning to be applied on a large scale in humans and in plant species such as maize. Application of this emerging technology to sugarcane is a major new development in sugarcane molecular genetics. SNP markers will be valuable in identifying polymorphisms that are associated with performance differences and consequently gene function. This project will act as a pilot to evaluate the suitability of SNPs as a marker not only as a general marker system, but for the study of specific genes in sugarcane.

1. ACHIEVEMENT OF OBJECTIVES

The proposal aimed to complete the validation within 12 months, commencing 01 July 2003 and ending 30 June 2004. The project commenced immediately EST sequence clusters were made available from Brazil in early June 2004. The bulk of the project was completed by late February 2004. The research proposal had the objective to evaluate the ability of the marker system to provide information on:

I. heterozygosity
II. origin of alleles
III. ability to identify single dose alleles
IV. ability to provide information beyond available current marker systems
V. protocols for SNP analysis in sugarcane using the pyrosequencer
VI. robustness of the pyrosequencing system used to detect SNPs
VII. development of database (not a formal objective)

I. HETEROZYGOSITY

Reports for diploid crop and animal species indicate that SNP markers are unable to reveal the same level of heterozygosity (average of 0.263 in elite maize germplasm) as an SSR marker potentially can. However, the levels of heterozygosity calculated for the markers developed in this project were remarkably high, with an average of 0.7285 ± 0.093. These figures however exclude markers that were monomorphic within the test population and genotypes used ranged across species. It is expected that SNP markers will be useful in fingerprinting and should therefore also be suited for genetic diversity studies as well.

Fingerprints derived from marker **SuSNP045-G511**
Click on **Variety** name for further information on that genotype

Variety	Allele Score
66N2008	A:68 G:32
Badila	A:57 G:43
IJ76-514	T:49 G:51
Korpi	A:50 G:50
Mandalay	A:100 G:0
POJ32878	A:100 G:0
Q117	A:75 G:26
Q124	A:65 G:35
Q165	T:75 G:25
Q96	A:100 G:0
R570	A:100 G:0
Tabango	A:100 G:0
Trojan	A:60 G:40

Fingerprints derived from marker **SuSNP010-C192**
Click on **Variety** name for further information on that genotype

Variety	Allele Score
66N2008	A:56 C:44
Badila	A:35 C:65
IJ76-514	A:15 C:85
Korpi	A:24 C:76
Mandalay	A:81 C:19
PJ2-1	A:19 C:81
PJ2-10	A:15 C:85
PJ2-2	A:17 C:83
PJ2-3	A:10 C:90
PJ2-4	A:16 C:84
PJ2-5	A:23 C:77
PJ2-6	A:15 C:85
PJ2-7	A:20 C:80
PJ2-8	A:16 C:84
PJ2-9	A:17 C:83
POJ32878	A:30 C:70
Q117	A:42 C:58
Q124	A:49 C:51
Q165	A:26 C:74
Q96	A:40 C:60
R570	A:39 C:61
Tabongo	A:100 C:0
Trojan	A:28 C:72

Example of fingerprints from two markers. The second marker (SuSNP010-C192) includes progeny of a mapping cross (PJ2). Whilst a single marker may not be able to distinguish all genotypes, a combination of markers will be able to fulfil this task.

II. ORIGIN OF ALLELES

A species specific marker has been identified, indicating the ability of these markers to identify allele origin.

Fingerprints derived from marker **SuSNP077-A2155**
Click on **Variety** name for further information on that genotype

Variety	Allele Score
66N2008	C:42 T:58
Badila	C:33 T:70
Black Innis	C:22 T:78
Djatiroto	C:100 T:0
Fiji 38	C:21 T:79
IJ76-514	C:42 T:58
IM76-250	C:55 T:45
IMP9068	C:100 T:0
IMP9819	C:100 T:0
IS76-147	C:38 T:62
Korpi	C:43 T:57
MQ30-2313	C:45 T:55
Mandalay	C:100 T:0
POJ2878	C:38 T:62
Q117	C:35 T:65
Q124	C:59 T:42
Q165	C:40 T:60
Q64	C:53 T:47
Q96	C:50 T:50
R570	C:35 T:65
SES147B	C:100 T:0
SES84/58	C:100 T:0
Tabongo	C:100 T:0
Tainan	C:100 T:0
Trojan	C:37 T:63

The marker, SuSNP077-A2155 has been able to identify the origin of the 'C' haplotype as having originated from *Saccharum spontaneum*. This was initially identified in the genotypes Tabongo and Mandalay, then later confirmed by testing the marker on additional *S. spontaneum* gentoypes Djatiroto, IMP9068, IMP9819, SES147B, SES 84/58 and Tainan; and *S. officinarum* genotypes Black Innis, Djatiroto, Fiji38, IM76-250 and MQ30-2313. In all cases, only the 'T' haplotype is present in *S. officinarum* and absent in *S. spontaneum*, making this marker likely to be *S. officinarum* specific.

III. IDENTIFICATION OF SINGLE DOSE ALLELES

Single dose markers have been found. However, only 2 such markers from 235 EST contigs examined were identified for the parents of the mapping population tested. This low number would indicate that whilst useful, this would be a costly exercise if the sole purpose for marker identification was to obtain markers for mapping.

Fingerprints derived from marker **SuSNP012-G2181**
Click on **Variety** name for further information on that genotype

Variety	Allele Score
66N2008	A:25 G:75
Badila	A:16 G:84
IJ76-514	A:0 G:100
Korpi	A:12 G:88
Mandalay	A:0 G:100
PJ2-1	A:12 G:88
PJ2-10	A:0 G:100
PJ2-2	A:10 G:90
PJ2-3	A:0 G:100
PJ2-4	A:10 G:90
PJ2-5	A:10 G:90
PJ2-6	A:0 G:100
PJ2-7	A:0 G:100
PJ2-8	A:0 G:100
PJ2-9	A:0 G:100
POJ32878	A:22 G:78
Q117	A:0 G:100
Q124	A:0 C:89 T:11 G:100
Q165	A:11 G:88
Q96	A:13 G:87
R570	A:10 G:90
R570	A:10 G:90
R570 self - A545	A:0 G:100
R570 self - A553	A:10 G:90
R570 self - A570	A:0 G:100
R570 self - A604	A:0 G:100
R570 self - A609	A:18 G:82
R570 self - A612	A:19 G:81
R570 self - A616	A:10 G:90
R570 self - A618	A:18 G:82
R570 self - A619	A:10 G:90
R570 self - A623	A:26 G:74
Tabongo	A:0 G:100
Trojan	A:0 G:100

Two markers were able to identify single dose alleles in the parents of the mapping population PJ2 (IJ76-514 x Q165). These were SuSNP061-C459 and SuSNP012-G2181. The progeny segregate 4:6 which is consistent with a 1:1 segregation of progeny from an outcross.

IV. ABILITY TO PROVIDE INFORMATION BEYOND AVAILABLE CURRENT MARKER SYSTEMS

This objective has not been directly shown in this validation project due to funds shortage and the number of markers currently available. However, the positive results obtained thus far indicate that these markers will be suited for:

i. *Fine mapping and candidate gene studies.*
Dense marker coverage is required in such studies and SNP markers are ideal for these.

ii. *High-density coverage*
Depending on the needs of the sugarcane industry, the number of SNP markers developed can greatly outnumber simple sequence repeats (SSRs) in the genome allowing for greater flexibility in marker selection and greater density than would otherwise be available.

iii. *High throughput*
SNP markers can be used to produce rapid, cost effective genotyping results. This is a clear benefit for high-throughput experiments designed to study familial linkage and to identify disease associated regions.

iv. *Greater genetic marker stability*
Mutations in SNPs are uncommon relative to SSR regions that are more likely to undergo deletions or insertions. SNP markers are therefore more robust from generation to generation, which is particularly important in linkage studies and studies on linkage disequilibrium.

v. *Ability to distinguish between haplotypes*
This has the potential to allow questions regarding 'active' alleles to be answered.

Essentially, it should be possible to carry out the same studies in sugarcane as are possible with diploid species.

V. PROTOCOLS USING PYROSEQUENCING FOR SNP ANALYSIS IN SUGARCANE

Protocols using the Pyrosequencer for SNP analysis in sugarcane have been developed. The main steps are in the PCR mix and cycling conditions. The focus is on the amount and quality of PCR product. Hence the use of a good quality Taq enzyme, number of cycles and $MgCl_2$ was the focus of optimisation.

Specific conditions are provided under section 2 – Methodology.

As previously noted, the second round PCR is a requirement only for the development of the marker and not necessary where primers are pre-biotinylated or where biotinylation is not required.

Issues with understanding and identifying reliable and unreliable runs in the pyrosequencer have been determined.

VI. ROBUSTNESS OF PYROSEQUENCING SYSTEM

The system for detection was tested for repeatability in terms of:

a. Technical repeatability
b. Biological repeatability

a. Technical Repeatability

A number of samples were selected for repeat scoring through separate PCR amplification steps and separate pyrosequencing runs. The results of these runs are tabulated below:

MARKER	GENOTYPE	REPEAT 1	REPEAT 2	**% variation**
SuSNP010-G339	R570	C:78.1%/T:21.9%	C:76.7%/T:23.3%	**1.4%**
SuSNP012-G2181	R570	G:89.4%/A:10.6%	G:90.5%/A:9.5%	**1.1%**
SuSNP036-C672	Q165	G:79.0%/T:21.0%	G:81.3%/T:18.7%	**2.3%**
SuSNP090-G1112	Q165	C:100.0%/T:0.0%	C:100.0%/T:0.0%	**0.0%**

b. Biological Repeatability

A second repeatability experiment to determine the consistency of results when different biological samples of the same genotype was carried out using 2 markers on 2 genotypes. The genotypes were collected from 6 separate locations across the state of Queensland in Australia and DNA extracted separately.

Marker: SuSNP068-T388

MARKER	SuSNP068-T388	SuSNP068-T388
GENOTYPE	Q117	Q124
LOCATION 1	A:71.8%/G:28.2%	A:73.3%/G:26.7%
LOCATION 2	A:74.5%/G:25.5%	A:78.0%/G:22.0%
LOCATION 3	A:72.0%/G:28.0%	A:69.6%/G:30.4%
LOCATION 4	A:71.5%/G:28.5%	A:74.0%/G:26.0%
LOCATION 5	A:73.6%/G:26.4%	A:74.6%/G:25.4%
LOCATION 6	A:72.7%/G:27.3%	A:79.0%/G:21.0%
% Variation	**0.5% to 3.0%**	**0.6% to 9.4%**

Marker: SuSNP077-A2155

MARKER	SuSNP077-A2155	SuSNP077-A2155
GENOTYPE	Q117	Q124
LOCATION 1	T:65.1%/C:34.9%	T:38.5%/C:61.5%
LOCATION 2	T:62.2%/C:37.8%	T:36.1%/C:63.9%
LOCATION 3	T:59.8%/C:40.2%	T:38.2%/C:61.8%
LOCATION 4	T:61.3%/C:38.7%	T:36.7%/C:63.3%
LOCATION 5	T:61.1%/C:38.9%	T:37.0%/C:63.0%
LOCATION 6	T:61.8%/C:38.2%	T:36.5%/C:63.5%
% Variation	**0.2% to 5.3%**	**0.3% to 2.4%**

Note that the high percentage variation using marker SuSNP068-T388 on Q124 is due to a poor pyrosequencing run on the genotype from Location 3.

The results from both the technical and biological repeatability runs indicate that variations will occur. In both cases, these variations can be attributed to the fact that PCR amplification does not produce amplification in perfect proportion in each run, although the small variations can be considered negligible. Queries into the Sequenome MALDI-TOF system indicate that variations of an average of 3% in diploid genomes are normal.

VII. DEVELOPMENT OF DATABASE

The development of a database did not form part of the objectives of the project, however it was nevertheless developed by the researcher. The database contains all the results of the project and provides a means in which to retrieve the data in a meaningful manner to maximise the information provided by the SNP markers. The database is located at:

http://www.scu.edu.au/research/cpcg/Sugar

and is password protected. ICSB members who have contributed to the project have been given access to the database. The 'EST sequence' provided in the database is in fact the consensus sequence of the contigs. The individual EST sequences will not be made available at this stage.

Summary of findings

Ease of **Development**

- The identification of potential SNPs and design of primers is relatively straightforward from clustered EST contigs.
- Confirmation is however a more time consuming process
- A set of 33 SNP markers were developed from 235 clustered EST contigs

Ease of High-Throughput **Scoring**

- Using the Pyrosequencer, high-throughput detection is straightforward and rapid.
- Alternative detection methods need to be investigated for laboratories without access to a Pyrosequencer.

Suitability for **Genotyping/Fingerprinting**

- 22 genotypes including commercial cultivars, ancestral species and mapping populations were genotyped
- Clear fingerprints were obtained. In some instances, a combination of markers will be required to distinguish between cultivars and/or ancestral species

Suitability for **Mapping**

- Two single dose markers were identified in the parents of the mapping population tested, indicating that there is potential for these markers to be used in mapping. A small number of progeny were screened with the markers producing expected segregation patterns. However, no attempt has been made to map these as yet.

Ease of **Allele Identification**

- Using the available techniques, it is currently not possible to identify individual alleles within any sugarcane genotype. It is possible to indicate the portion of each allele present in the genotype.

Species specific markers

- Markers have been identified that are potentially species specific (S. spontaneum)
- Further verification will be required.

2. Methodology

For each marker, one pair of PCR primers and one sequencing primer (or SNP detection primer) was required. The sequencing primer was preferentially designed in the forward direction and complimentary to the reverse strand. Where conditions did not suit, the sequencing primer was designed in the reverse direction and complimentary to the forward sequence.

Only the complimentary strand to the sequencing primer must be present in the pyrosequencer, hence a means of separating and selecting the required strand is required. This (as part of the preparatory stage for pyrosequencing) was achieved by tagging the desired strand with a biotin label and capturing the strand with a magnet.

The cost of biotinylating each forward or reverse PCR primer is prohibitive in a developmental phase. Hence, a method to use a universal biotin labelled 11-mer for attachment to the relevant strand was developed [**Pacey-Miller T and Henry R** (2003). SNP detection in plants using a single stranded pyrosequencing protocol with a universal biotinylated primer. Analytical Biochemistry 317: 165-170.]. In brief, to select the reverse PCR fragment, the reverse PCR primer was synthesised with the additional tag sequence to the primer: 3' GCCCCCGCCCG-GAATTAGTGAAGGCGAAGA 5'. In a second round PCR using the first round PCR product as template, the biotin label attached to the generic sequence was used as the reverse primer: biotin-CGGGGGCGGGC.

The identification of alternative SNP scoring systems not requiring primer biotinylation will be investigate in a subsequent project.
The single stranded template (PCR product) is added together with the SNP detection primer to the Pyrosequencer. Sequence extension then determines the proportional presence of each SNP base.

Criteria for designing PCR Primers

1. Product length between 100 bp and 200 bp
2. PCR primer length 21 ± 3 bp
3. T_m between 45 ^oC - 55 ^oC
4. No 3'-end complementarity greater than 2 bp
5. GC content: Min 40%, Max 60%

Criteria for designing Sequencing Primers (ie SNP detection primers)

1. 0-5 bp from SNP
2. No 3'-end complementarity greater than 2 bp
3. ~50^oC T_m (range ~(43 ^oC to 53 ^oC)
4. Primer length 13 bp – 19 bp
5. Proceeding base not identical to SNP bases

Primers were designed using the software, Primer Premier v 5.00 by PREMIER Biosoft International. Synthesis was done at Proligo, Lismore, Australia.

3. Nomenclature of Primers

PCR primers

PCR primers were labelled according to the Contiguous sequences they were derived from.

For example, primers from Contig 11 are named:

Contig 11.1-F and Contig 11.1-R

The suffix number indicates that one or more PCR primers may be designed to a Contig. F and R represent Forward and Reverse respectively.

Sequencing or SNP detection Primers

SNP detectiong primers were labelled according to the Contiguous sequence on which they were derived, the consensus SNP and its position.

For example, the marker **SuSNP077-G2511** would refer to

SuSNP	:	A prefix to indicate 'Sugarcane SNP'
077	:	The contiguous sequence from which the SNP was derived.
G	:	The SNP base in the consensus sequence. The variant base is found as part of the information in the database.
2511	:	The base position of the SNP on the consensus sequence.

In instances where the sequencing primer is designed in the reverse direction, the consensus SNP base and its base position remain as read in the forward direction.

PCR primer sequences are listed in **APPENDIX II,** and respective SNP detection (sequencing) primers listed in **APPENDIX I**.

3. Issues with Development

i. Attrition

The number of potential SNPs identified and the number of useable markers is disparate. From 235 EST sequences analysed, approximately 200 potential SNP markers were identified. However, only 86 SNPs could have PCR primers and SNP detection primers designed.

ii. PCR optimisation

The majority of primers designed were robust enough to amplify clean bands at each PCR. However, a number required optimisation that was time consuming. Amongst these, the easy option was abandonment. However, a small number appeared to be SNPs worth persisting in development. Cases where this would be, include SNPs identified in genes/ESTs of interest.

iii. Identified in ESTs but verified in genomic DNA

Whilst ESTs were used to identify potential SNPs, genomic DNA was used to validate their presence. Introns were invariably present in most cases which caused two problems. The first causes the PCR fragment size to be beyond the optimum recommended for use in the pyrosequencer (ie less than or equal to 200 bp). Whilst fragments greater than 500 bp were abandoned, fragments between 200 and 500 bp were still used. In several cases, the altered sequence affected the pyrosequencing reads casting doubt on the validity of the results.

iv. Access to genotypes used in EST development

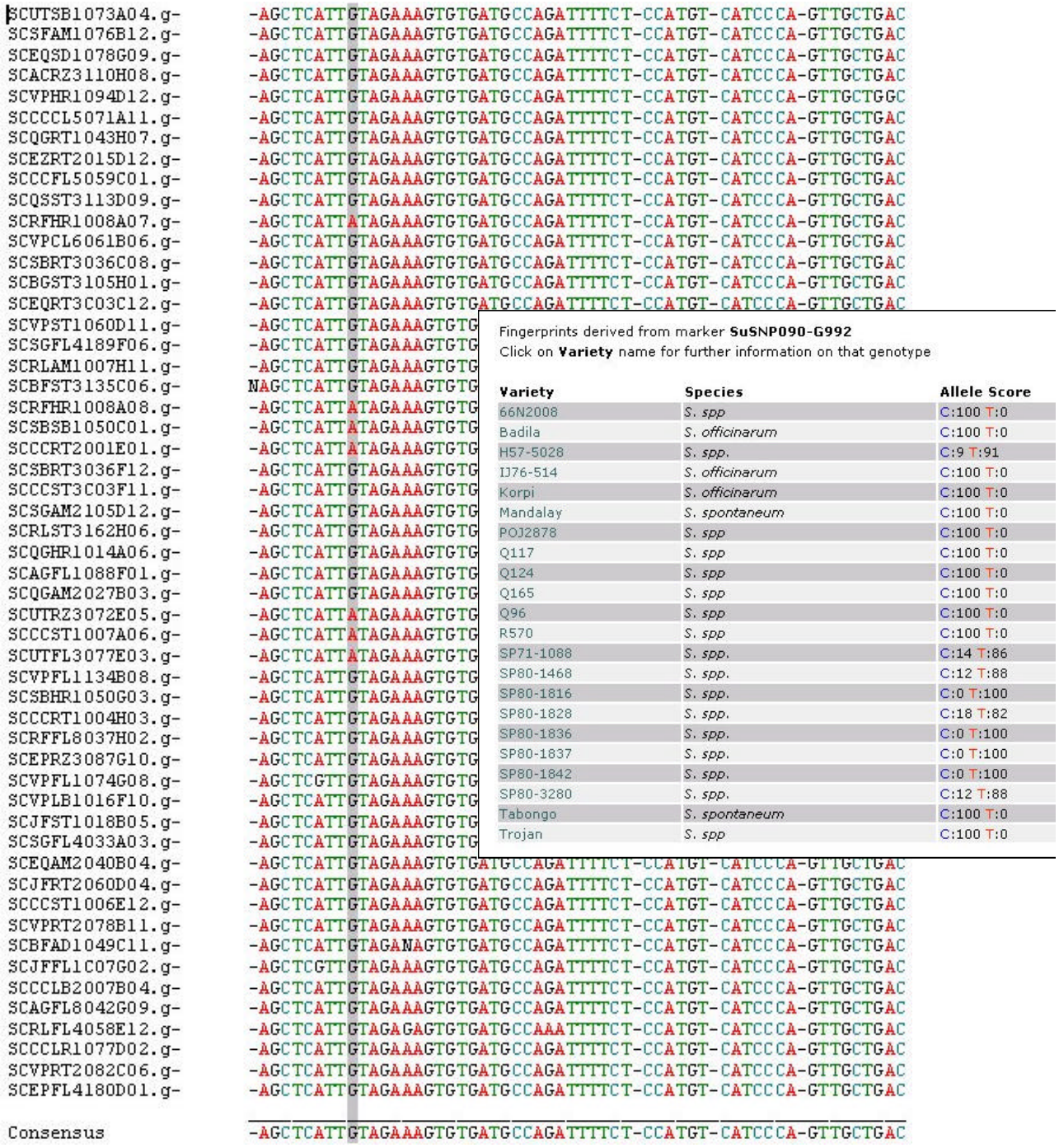
In 5 instances, EST clusters indicate the presence of a potentially genuine SNP. However pyrosequencing did not detect a SNP at the loci in the genotypes tested. Without the genotypes from which the ESTs were derived, it was not possible to validate the potential SNP. Further investigation into this issue showed that in all 5 cases, the variant base (ie the variation to the consensus sequence) was derived from the Brazilian variety, SP80-3280. The variant base was present in only 2 other Brazilian varieties used in the development of the ESTs. Confirmation on the validity of these markers was made after DNA was sourced from a select set of Brazilian sugarcane genotypes.

The example on the following page shows a marker designed to rare alleles found to our knowledge, only in the Brazilian genotypes. The variant base was not detected in any of the core genotype samples, but was found in 4 of 8 Brazilian sugarcane genotypes (SP) tested.

SuSNP090-C992

53 ESTs : 7A/46G

All ESTs that contained the A base at the SNP locus also contained single nucleotide polymorphisms at other loci both in the upstream and downstream direction (not shown) indicating the A variant to be a genuine SNP. In this example, the variant A SNP is derived from only 2 genotypes (one of which is SP80-3280). Score results show polymorphisms only in SP genotypes.



4. Summary and Recommendation

i. Fingerprinting is possible. Suitability for genetic diversity and/or phylogenetic analysis in sugarcane requires confirmation.
ii. Results are repeatable with a small percentage variation (0 to 3%) when using the pyrosequencer.
iii. A species specific (*S. officinarum*) marker has been identified, indicating the potential for additional such markers to be found.
iv. Single dose markers identified for the mapping population PJ2 and tested progeny show a consistent segregation pattern.

The large number of EST sequences recently lodged in GenBank by FAPESP will allow the rapid selection of markers that will prove robust and useful, and avoid the need to optimise 'recalcitrant' markers, thus reducing any wastage of time and finances.

Where (ancestral) species specific markers are developed, it may be possible to use these markers to identify the portion of a genotype that has been derived from that particular ancestral species. Where QTL markers are developed, it may be possible to predict the level of trait expression against allele dose.

Recommendation

For the purpose of straight fingerprinting, SSR markers provide the necessary level of heterozygosity and are easy to use, particularly if analysed with automated systems. Fingerprinting with SNPs can also provide a high level of discrimination but will require more 'sophisticated' detection systems.

However, for the reasons covered in this report, including fine genome mapping, marker assisted selection, high throughput genotyping, greater marker stability and the ability to distinguish between haplotypes, it would be imprudent for the sugarcane community not to pursue the development of a SNP marker system for sugarcane. SNP markers can be relatively easily designed to target genes of interest, whilst there is a lack of SSR markers that can be used to target and analyse genes of interest.

5. Publications

Cordeiro GM & Henry RJ (2004). Validating single nucleotide polymorphisms as a new marker system for understanding the Saccharum genome. *Plant & Animal Genomes Conference XII, San Diego.* Abstract W122.

SNP NAME	EST ID	BASE	POSITION	SEQUENCING PRIMER	STRAND	TM	DATE CREATED	LENGTH	EXPECTED PRIMED SEQUENCE
SuSNP060-G228	CONTIG 60.1	G|A	228	CGGGCGAGCTGGA	F	50.5	16/06/2003	15	T**G|A**TCCTTGGGCA
SuSNP126-T1626	CONTIG 126	T|C	1626	GGTTGAGTACTTCGG	F	37.5	30/10/2003	15	**T|C**GAGCAATTGTCT
SuSNP141-G470	CONTIG 141.1	G|A	470	GACCGGTTAGCTCGAC	R	47.4	3/11/2003	16	**C|T**GCCCAACCGAT
SuSNP010-C192	CONTIG 10.1	C|A	192	GGATGTAAACTTGGTAAC	F	41.1	16/06/2003	18	**C|A**GCCTTGGTGC
SuSNP010-G339	CONTIG 10.2	C|T	339	AACTCCTTCATCAACG	R	41.0	16/06/2003	16	A**C|T**ATCTTCGAGA
SuSNP012-C2081	CONTIG 12.1	C|T	2081	GGCGACAGAAATCCTA	F	45.0	16/06/2003	16	G**C|T**GGCAAGTCT
SuSNP012-C3295	CONTIG 12.4	C|A	3295	ATCCTGACTGTGCCTGG	F	50.2	16/06/2003	17	**C|A**GCTGCCACAT
SuSNP012-G2181	CONTIG 12.2	G|A	2181	ACCCAGAAGAAGCAAC	F	43.5	16/06/2003	16	C**G|A**CCTGCATTGC
SuSNP214-A2882	CONTIG 214.2	A|G	2882	TGGGTGGCACAGGTTC	R	51.7	5/12/2003	16	**C|T**GCCATTGAGAAGA
SuSNP214-G2048	CONTIG 214.1	G|A	2048	CTGCGGTTGCATATCCAC	R	54.9	3/12/2003	18	**T|C**AGGGTTCGACAGG
SuSNP110-G424	CONTIG 110.2	G|T	423	CCTTGGTCATCTTCCC	F	46.5	22/10/2003	16	A**G|T**ACTCGACAGCT
SuSNP141-G506	CONTIG 141.2	G|C	506	CAACAGAGAAACCCATC	R	44.0	4/11/2003	17	**C|G**TCGCCAACGGAGT
SuSNP166-A490	CONTIG 166.2	A|G	490	GCTGTGTTTGTTGACCT	R	44.5	11/11/2003	17	**T|C**GAGCCCACTGTC
SuSNP166-G319	CONTIG 166.1	G|A	319	TGGAGACCAGTGCAGTT	F	47.7	12/11/2003	17	**G|A**TCTGCAAGCTTCC
SuSNP028-C517	CONTIG 28.2	G|A	517	GTCTATCTTCTCCCTGC	R	42.6	16/06/2003	17	T**G|A**TGGGTGAAGA
SuSNP036-C672	CONTIG 36.1	G|T	672	AGGCAGCCATGTTTTCA	R	43.6	16/06/2003	17	**G|T**ATCCAAACAT
SuSNP175-A2508	CONTIG 175.3	A|G	2508	AAGTGGTTTCAGTGTTAG	F	41.1	14/11/2003	18	T**A|G**TCCTTATATGAAG
SuSNP175-T1280	CONTIG 175.1	T|C	1280	CTCCATCCTTCCACTA	F	40.8	15/11/2003	16	**T|C**GATTTGCTTCTCTT
SuSNP184-T1023	CONTIG 184.1	T|C	1023	CACATCAAGATTTCAA	R	35.8	21/11/2003	16	**A|G**CTTCGCATA
SuSNP045-G511	CONTIG 45.1	G|A	511	TCGACCCTCCTATCCAGA	F	52.6	16/06/2003	18	C**G|A**CTGTACTTCC
SuSNP224-G193	CONTIG 224.1	G|A	193	AGGGCAAAGCTAGTAAGG	F	49.6	26/11/2003	18	**G|A**CGATATCGCCGGC
SuSNP053-T421	CONTIG 53.1	T|G	421	CGCGCTCATGGATGG	F	53.8	16/06/2003	15	**T|G**CTGCT
SuSNP192-C533	CONTIG 192.1	C|T	533	GGCTTCCTCGTCTTCA	F	47.6	16/06/2003	16	A**C|T**GCTGTTGGTGGAGG
SuSNP056-G1534	CONTIG 56.1	G|A	1534	CATCTGAACCATCTGCC	F	47.8	16/06/2003	17	**G|A**CCTGAGGTTC
SuSNP061-C459	CONTIG 61.1	C|T	459	GCCCGGTGTTGAGGT	F	50.7	16/06/2003	15	**C|T**GAGGTGACCA
SuSNP068-T388	CONTIG 68.1	A|G	388	AGAGGCGGCATCACTG	R	51.5	16/06/2003	16	C**A|G**CCAGCTGGTT
SuSNP077-A2155	CONTIG 77.2	T|C	2155	TACCAGCTTGGAAAACTCA	R	51.2	16/06/2003	19	A**T|C**GAGGCCACTT
SuSNP077-C903	CONTIG 77.1	C|T	903	GTGGATTTCAAGGAGAGTG	F	49.2	16/06/2003	19	A**C|T**ATCAAGTGGG
SuSNP081-A446	CONTIG 81.1	A|G	446	GGTGGAAGGCTCATCA	F	47.8	16/06/2003	16	C**A|G**TCGCAGGGAA
SuSNP087-C819	CONTIG 87.2	C|T	819	CCCCTCCAGGATGTGT	F	49.5	16/06/2003	16	A**C|T**AAGATTTGGT
SuSNP090-G1112	CONTIG 90.3	C|T	1112	GGGATGGGACAGAAGG	R	49.3	16/06/2003	16	A**C|T**GCATATGTTG
SuSNP090-G632	CONTIG 90.1	G|A	632	GCCAGCTTCTCCTTGA	F	47.9	16/06/2003	16	T**G|A**TCCCTTACAA
SuSNP090-G992	CONTIG 90.2	C|T	992	ATCTGGCATCACACTTTCT	R	49.2	16/06/2003	19	A**C|T**AATGAGCTTC

TEMPLATE PRIMER SEQUENCES

Appendix II

SNP_ID	PRIMER_NAME	LENGTH	PRODUCT_SIZE	SEQUENCES	TM	STRAND
SuSNP010-C192	Contig 10.1F	18	160	GGAAACGAAACAGAACCG	53.3	F
SuSNP010-C192	Contig 10.1R	18	160	AGAAGCCCACCATCACCT	54.3	R
SuSNP010-G339	Contig 10.2F	18	113	GTGGGCTTCTTGTTGTAG	47.6	F
SuSNP010-G339	Contig 10.2R	18	113	GACATTGGTATCTCGTCC	46.5	R
SuSNP012-C2081	Contig 12.1F	18	154	ATTGCACTCTTGGCTCTT	49.7	F
SuSNP012-C2081	Contig 12.1R	20	154	CATCATTCGTGTCATTTGTC	51.1	R
SuSNP012-G2181	Contig 12.2F	18	116	CTGTGGCAACCAGCAAAT	54.5	F
SuSNP012-G2181	Contig 12.2R	22	116	GAGGCAGACATAACATAGGAGT	53.1	R
SuSNP012-C3295	Contig 12.4F	18	165	AGCACAGGTGGTGAATGG	53.6	F
SuSNP012-C3295	Contig 12.4R	20	165	CTAAACTCAGCAGGAAAAGA	50	R
SuSNP028-C517	Contig 28.1F	19	187	CAGGTATTTACACGAGGGG	52.8	F
SuSNP028-C517	Contig 28.1R	18	187	ATCGTCATCGTCTGGTCC	52.4	R
SuSNP036-C672	Contig 36.1F	18	149	GCTGCAACAAAACCTGAA	52.1	F
SuSNP036-C672	Contig 36.1R	18	149	CCAAAGGAAGAGGCACAG	53.2	R
SuSNP045-G511	Contig 45.1F	18	189	CATCATACTCCTCCTTCG	47.2	F
SuSNP045-G511	Contig 45.1R	18	189	GAGGTTCGACTATGTTCC	46.1	R
SuSNP053-T421	Contig 53.1F	18	185	GAGGGGAAGAAGCTGTCT	50.7	F
SuSNP053-T421	Contig 53.1R	19	185	GAATTAGTGAAGGCGAAGA	50.1	R
SuSNP056-G1534	Contig 56.1F	18	100	GAGAAGGACGTGGTGTAG	46.8	F
SuSNP056-G1534	Contig 56.1R	18	100	TCAGAAATACCAGGAAGC	47.3	R
SuSNP060-G228	Contig 60.1F	18	195	ACCAAACTACAGGACCCA	49.7	F
SuSNP060-G228	Contig 60.1R	18	195	CTCTTCGAGGACACCAAC	49.6	R
SuSNP061-C459	Contig 61.1F	18	186	ATGGGATCGGTTTGAGTT	53.3	F
SuSNP061-C459	Contig 61.1R	20	186	GATAGTTGATGACATTGCTG	53.7	R
SuSNP068-T388	Contig 68.1F	18	170	GCTTGGTGACGCTTCTTC	51.8	F
SuSNP068-T388	Contig 68.1R	18	170	CTTGTTGTCCGGTCCTTG	48	R
SuSNP077-C903	Contig 77.1F	20	177	ATGGAATGATAAGAAGACCC	50.6	F
SuSNP077-C903	Contig 77.1R	19	177	ATCAATAGATAGCTGTCCC	46	R
SuSNP077-A2155	Contig 77.2F	18	195	TTGCCACGAAGAAAGAAG	51.5	F
SuSNP077-A2155	Contig 77.2R	19	195	CACAGCCCAAAACAGAAAA	54.3	R
SuSNP081-A446	Contig 81.1F	18	190	GCCATTTCACGCAACATC	54.8	F
SuSNP081-A446	Contig 81.1R	18	190	TAGCCATCTCAACCATC	48.8	R
SuSNP087-C819	Contig 87.2F	18	163	CCTGCTTGAGGCTCTTGA	53.8	F
SuSNP087-C819	Contig 87.2R	20	163	ACCGAAGGTGACAACCATA	54.4	R
SuSNP090-G632	Contig 90.1F	18	135	GTTTCCAACTCTTGCTCA	47.8	F

TEMPLATE PRIMER SEQUENCES

Appendix II

SNP_ID	PRIMER_NAME	LENGTH	PRODUCT_SIZE	SEQUENCES	TM	STRAND
SuSNP090-G632	Contig 90.1R	18	135	CCTCACAGACTCCCTCAT	48.5	R
SuSNP090-G992	Contig 90.2F	18	151	CGAACATAATCTGGGTCA	48.5	F
SuSNP090-G992	Contig 90.2R	18	151	ATTGTCAGCAACTGGGAT	49.7	R
SuSNP090-G1112	Contig 90.3F	18	127	ATTGGGTACTTGAGCGTG	50.9	F
SuSNP090-G1112	Contig 90.3R	18	127	CTGTTTTCCCTAGCATTG	48.3	R
SuSNP110-G424	Contig 110.2F	20	186	GTGACATTTGCAGATCCGTG	56.7	F
SuSNP110-G424	Contig 110.2R	20	186	TGCTTCAATCTTTGCGTGGA	60	R
SuSNP126-T1626	Contig 126.1F	20	143	CATCAGTGCCATCAAGGAAG	56	F
SuSNP126-T1626	Contig 126.1R	20	143	CGGTGAATGCAAAACCAGAC	58.7	R
SuSNP141-G470	Contig 141.1F	20	161	CTTCAGGACATCCAACCTCG	57.6	F
SuSNP141-G470	Contig 141.1R	20	161	CAACGGAATCACCACCACTC	58.1	R
SuSNP141-G506	Contig 141.1F	20	161	CTTCAGGACATCCAACCTCG	57.6	F
SuSNP141-G506	Contig 141.1R	20	161	CAACGGAATCACCACCACTC	58.1	R
SuSNP166-G319	Contig 166.1F	20	197	ATCCCGTTCCTCCACCAACA	62.1	F
SuSNP166-G319	Contig 166.1R	20	197	CCTACCGCCAGCTCTTCCAC	62	R
SuSNP166-A490	Contig 166.2F	20	159	GAAGAGCTGGCGGTAGGTGC	62.2	F
SuSNP166-A490	Contig 166.2R	20	159	CAGATGCCCGGTGACAAGAC	60.8	R
SuSNP175-T1280	Contig 175.1F	20	174	GCCTTCATACCAGCAACGAT	57.6	F
SuSNP175-T1280	Contig 175.1R	20	174	TGATGGAGGGATACCAATAA	52.7	R
SuSNP175-A2508	Contig 175.3F	20	168	TAAGCTATGCAACTACTGGC	51.2	F
SuSNP175-A2508	Contig 175.3R	22	168	CAGGTATAGCACTCCTTTAATC	51.3	R
SuSNP184-T1023	Contig 184.1F	19	165	AATGCTATGCTTCAAGTCC	49.9	F
SuSNP184-T1023	Contig 184.1R	20	165	TAGTTTCAATCATCTACCCC	49.2	R
SuSNP214-A2882	Contig 214.2F	22	153	AAGGTTGAAGACTACCGGAACA	58.3	F
SuSNP214-A2882	Contig 214.2R	20	153	CTCCCTCAGCGTGATGTTGC	61.2	R
SuSNP224-G193	Contig 224.1F	21	148	CACACGGTGATGGGAGTAGAT	57	F
SuSNP224-G193	Contig 224.1R	21	148	CCTGCAACAACCTGGTGGAAT	62.1	R