2007

Predicting progeny performance and genome analysis in sugarcane (Saccharum spp)

Sreedhar Alwala
Louisiana State University and Agricultural and Mechanical College

PREDICTING PROGENY PERFORMANCE AND GENOME ANALYSIS IN SUGARCANE (*SACCHARUM* SPP)

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The School of Plant, Environmental and Soil Sciences

By
Sreedhar Alwala
B. Sc (Agri), Acharya N. G. Ranga Agricultural University, 1998
M. Sc (Plant Breeding and Genetics), Tamil Nadu Agricultural University, 2000
August, 2007

Dedicated to my mother Smt. Bharathi

ACKNOWLEDGEMENTS

I am earnestly thankful to Dr. Collins Kimbeng who is more a friend and a philosopher than a major professor. His advice, support and patience throughout the program is invaluable. The freedom he has provided during the course of study is commendable. The healthy discussions with him have shaped my critical thinking and my career as well. I will always be grateful to him for all his help during the times I needed most. I have really enjoyed working with him.

I would like to extend my sincere gratitude to my committee members Dr. Manjit Kang for giving me an opportunity to do research with him, Dr. Kenneth Gravois for immensely helping me with the field trial at St. Gabriel Research station, Dr. Prasanta Subudhi for allowing me to do molecular marker work in his lab during the initial period of study and Dr. John Veremis for providing me with the mapping population and Dr. Charlie Johnson for accepting to serve in my committee at such a short notice. I also thank all my committee members for their constructive suggestions while preparing this dissertation.

I would also like to thank my lab mates: Jie , Suman and Marvellous for helping me in collecting the field data. I am also thankful to Dr. Baisakh and Ashok for helping during the course of study. The help rendered by St. Gabriel Research Station field crew is highly appreciated. My heartfelt thanks are also extended to Chandrasekhar, Vikrant, Sushil, Pramod, Srinivas and Srinivas without whom I would have missed the fun and wonderful life in Baton Rouge.

I am grateful to my parents Mr. Rajalakshman and Mrs. Bharathi and all my family members: Padmaja, Anand Rao, Venkatesh, Rani, my wife Deepthi, Niraj, Vaishnavi and Vinayak and for their unconditional love and affection. Special thanks to

my childhood friends Kiran and Srinath for their support.

The financial support from Louisiana Board of Regents, American Sugarcane League and LSU Graduate School Dissertation fellowship is highly appreciated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Genetic diversity/similarity (GS) was estimated among nine sugarcane parental genotypes using target region amplification polymorphism (TRAP), amplified fragment length polymorphism (AFLP) markers and coefficient of parentage (*f*). Its relationship to progeny performance was assessed among five bi-parental crosses involving the nine parents. Phenotypic data for stalk-height, -count, -diameter, cane yield and theoretical recoverable sugars (TRS) were collected from which genetic parameters (family means, genetic variances, average mid parental heterosis, percent heterotic clones and mid parental values) were calculated. For TRS, families with higher means and variances had a greater proportion of heterotic clones. AFLP-GS was found to be a good predictor of genetic parameters for most of the traits and TRAP-GS (and *f*-GS) could be used as a good predictor for TRS.

Framework linkage maps of *Saccharum officinarum* 'Louisiana Striped' and *S. spontaneum* 'SES 147B' were constructed using AFLP, sequence related amplified polymorphism (SRAP) and TRAP markers: the *S. officinarum* map comprised of 146 linked markers spanning 49 linkage groups (LG) and the *S. spontaneum* map comprised of 121 linked markers spanning 45 LG. Compared to AFLP, SRAP and TRAP markers appear less efficient for linkage mapping in sugarcane, a complex polyploid. The S*accharum* interspecific F_1 mapping population was evaluated for Brix(B), pol(P) and sucrose(S) at the early(E) and late(L) plant growing seasons in two years, 2004(04) and 2005(05). Conventional quantitative trait loci (QTL) analysis identified markers associated with these traits. In *S. officinarum*, 50 QTLs were identified with LOD scores ranging from 2.51 to 7.64, explaining from 15.9% (04LP) to 47.8% (04EB) of phenotypic variation. In *S. spontaneum*, 26 QTLs were identified with LOD scores ranging from

2.69 to 7.51, explaining from 6.5% (04LP) to 43.5% (04LB) of phenotypic variation. Thirty-four digenic interactions were observed in *S. officinarum* and four in *S. spontaneum*. Several SRAP and TRAP markers were found to be associated with traits indicating their potential usefulness in QTL tagging. A non-parametric approach, discriminant analysis (DA), also identified several markers which were either similar or localized to the same genomic regions as identified by QTL analysis validating DA as a viable option to identify marker-trait associations. Markers identified in this study in both parents would serve in marker assisted introgression breeding for sugarcane improvement.

CHAPTER 1 GENERAL INTRODUCTION

Sugarcane (2*n* = 128 – 140) belongs to the genus *Saccharum* of the Poaceae family. The genus *Saccharum* is comprised of six species, namely *S. officinarum* L.*, S. barberi* Jesw., *S. sinense* Roxb.*, S. spontaneum* L.*, S. robustum* Brandes and Jeswiet ex Grassl, and *S. edule* Hassk. (Brandes, 1958). Before the 19th century, *S. officinarum* (2*n* = 80) represented the cultivated sugarcane and is characterised by high sugar content, thick stalks, low fiber and low disease resistance. In the early 19th century, after witnessing the devastating effects of sereh disease, interspecific hybridizations were initiated between the cultivated *S. officinarum* and the wild *S. spontaneum*. The *S. spontaneum* species (2*n* = 64 to 128) is characterised by low sugar content, thin stalks, high fiber, high ratooning ability and high resistance to biotic and abiotic stresses. Therefore, to minimize the negative effects of *S. spontaneum* and to retain the high sucrose producing ability of *S. officinarum*, a series of backcrosses were made between the interspecific hybrids and the *S. officinarum* parent. This process was termed as 'nobilization' (Sreenivasan et al., 1987) and provided a major breakthrough in sugarcane improvement programs in the form of improved productivity, high disease resistance and high ratooning ability.

Although nobilization was highly successful, limited progress has been made to improve sugar content in the past few decades in most sugarcane breeding programs. The major reason attributed for the limited progress is the narrow genetic base of sugarcane cultivars (Deren, 1995; Arro, 2005). Very few clones were used in the initial interspecific hybridizations (Arcenueaux, 1965) and most of the modern day sugarcane cultivars trace back to those few parents used during nobilization. Currently, most sugarcane improvement programs follow a similar trend of making crosses among the

related parental genotypes followed by clonal selections. Therefore, the chances of creating new genetic variation are very meager. In addition, the long selection cycle of 12 – 15 years and clonal propagation of sugarcane indicate that the development of sugarcane cultivars has not involved many meiotic events. Hence, there is a growing concern among the sugarcane breeders regarding the narrow genetic diversity among the sugarcane cultivars.

A thorough understanding of the genetic diversity among the sugarcane genotypes will facilitate the development of improved cultivars. The information on the genetic relationships would be useful in utilizing and managing the sugarcane genotypes and the gene pool in breeding programs. Based on the genetic diversity knowledge, crosses could be designed among the divergent parents to maximize heterosis and resources could be allocated to the most promising crosses. As the sugarcane improvement programs thrived on few crosses followed by recurrent selections, a plateau could have been reached for improving sugar content. Currently, the genome of modern sugarcane cultivars is comprised mainly of *S. officinarum* (up to 80%) genome (D'Hont et al., 1996) with *S. spontaneum* accounting for most of the diversity within cultivars (Jannoo et al., 1999). This may suggest that most of the favourable alleles for sugar content have already been accumulated in cultivars. Therefore, exploring wild germplasm for novel sources of genes could be useful in sugarcane breeding programs. The newly identified genes/alleles from wild germplasm would help in widening the genetic base via introgression breeding.

An advance in molecular biology through the development of molecular marker technology has been pivotal in deducing inferences about genetic diversity and inter-relationships among the sugarcane genotypes at the DNA level (D'Hont et al., 1996;

Jannoo et al., 1999; Lima et al., 2002; Cordeiro et al., 2003). In addition, molecular markers have played a vital role in tracking favorable alleles from wild species as well as ascertaining their introgression into the cultivated background (Edmé et al., 2006). In sugarcane, several molecular marker based genetic linkage maps have also been constructed in the progenitor species such as *S. officinarum* and *S. spontaneum* (da Silva et al., 1993; Al-Janabi et al., 1993; Mudge et al., 1996; Ming et al., 1998; Edmé et al., 2006) and in commercial cultivars (Hoarau et al., 2001; Aitken et al., 2005; Reffay et al., 2005).

Although, the current molecular marker tools, such as RFLP, RAPD, AFLP, gSSR have been very useful in studying the patterns of genetic diversity and dissecting the genome, the polymorphisms generated by these marker techniques are randomly distributed across the genome. In addition, the marker-QTL associations are mostly distal linkages and the transferability of QTLs between populations still remains a major problem. Therefore, the polymorphisms which are directly associated with the genes/QTLs would be of immediate interest to the breeder. With the advent of modern bioinformatic tools and increasing access to vast numbers of EST sequences, it is now possible to explore new opportunities in sugarcane molecular marker research.

Recently, two new PCR-based markers namely, SRAP (sequence related amplified polymorphism) and TRAP (target region amplification polymorphism) which amplify intragenic polymorphism have been reported. In the SRAP technique, the primers are arbitrarily designed with an AT- and GC-rich motif to anneal to intron and exons, respectively (Li and Quiros, 2001), whereas in the TRAP technique, the forward or fixed primer is designed using gene/EST sequence information and the reverse primer is similar to a SRAP primer (Hu and Vick, 2003). The amplifications generated by these

two marker techniques supposedly reflect the polymorphisms within functional genes. The objectives of this research were

1. To predict the cross progeny performance based on TRAP-, AFLP- and coefficient of parentage (*f*) estimated genetic similarity among sugarcane genotypes
    a. To evaluate the potential of TRAP markers to assess the genetic diversity among sugarcane germplasm collections
    b. To compare TRAP marker based genetic similarity with AFLP- and *f*-based genetic similarity among nine elite sugarcane parental genotypes
    c. To predict the progeny performance of five bi-parental crosses involving nine parents based on molecular marker and *f*-based genetic relationships.
2. To dissect the genome of progenitor species of sugarcane namely *S. officinarum* and *S. spontaneum* in order to identify novel alleles for introgression breeding
    a. Genetic linkage mapping of *S. officinarum* 'Louisiana Striped' and *S. spontaneum* 'SES 147B' using AFLP, TRAP and SRAP markers
    b. Identification of molecular markers associated with sugar related traits in a *Saccharum* interspecific cross using conventional quantitative trait loci (QTL) analysis.

1.1 References

Aitken, K. S., P.A. Jackson and C. L. McIntyre. 2005. A combination of AFLP and SSR markers provide extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor. Appl. Genet. 110: 789-801

Al-Janabi, S. M., R. J. Honeycutt, M. McClelland and B. W. S. Sobral. 1993. A genetic linkage map pf Saccharum spontaneum L. 'SES 208'. Genetics 134: 1249-1260

Arcenueaux, G. 1965. Cultivated Sugarcane of the world and their botanical derivation, p. 844-854 Proc Int Soc Sug Tech, Vol. 12.

Arro, J. A. 2005. Genetic diversity among sugarcane clones using Target Region Amplification Polymorphism (TRAP) markers and pedigree relationships. Masters Thesis submitted to Louisiana State University, Baton Rouge, LA.

Brandes, E. W. 1958. Origin, classification and characteristics. In: E Artschwager and EW Brandes, Sugarcane (*S. officinarum* L.). USDA Agri. Handbook. 122 pp 1-35, 260-262

Cordeiro, G.M., Y. B. Pan and R. J. Henry. 2003. Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. Plant Science 165:181-189.

da Silva, J. A. G., M. E. Sorrells, W. Burnquist and S. D. Tanksley. 1993. RFLP linkage map and genome analysis of *Saccharum spontaneum.* Genome 36: 782–791.

Deren, C.W. 1995. Genetic base of U.S. mainland sugarcane. Crop Sci. 35:1195-1199.

D'Hont, A., L. Grivet, P. Feldmann, S. Rao, N. Berding and J.C. Glaszmann. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp) by molecular cytogenetics. Mol Gen Genet 250:405-413.

Edmé´, S. J., N. G. Glynn and J. C. Comstock. 2006. Genetic segregation of microsatellite markers in *Saccharum officinarum* and *S. spontaneum.* Heredity 97: 366-375

Hoarau, J.Y., B. Offmann, A. D'Hont, A. M. Risterucci, D. Roques, J. C. Glaszmann and L. Grivet. 2001. Genetic dissection of a modern sugarcane cultivar (*Saccharum spp*.). I. Genome mapping with AFLP markers. Theor. Appl. Genet. 103:84–97

Hu, J.G., and B. A. Vick. 2003. Target region amplification polymorphism: A novel marker technique for plant genotyping. Plant Mol. Biol. Rep. 21:289–294.

Jannoo, N., L. Grivet, M. Seguin, F. Paulet, R. Domaingue, P.S. Rao, A. Dookun, A. D'Hont, and J.C. Glaszmann. 1999. Molecular investigation of the genetic base of sugarcane cultivars. Theor Appl Genet 99:171-184.

Li, G., and C. F. Quiros. 2001. Sequence related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: Its application to mapping and gene tagging in Brassica. Theor. Appl. Genet. 103:455–461.

Lima, M.L.A., A. A. F. Garcia, K. M. Oliveira, S. Matsuoka, H. Arizono, C. L. de Souza and A.P. de Souza. 2002. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum spp*.). Theor Appl Genet 104:30-38.

Ming, R., S. C. Liu, Y. R. Lin, J. A. G. da Silva, W. Wilson, D. Braga, A. van Deynze, T. F. Wenslaff, K. K. Wu, P. H. Moore, W. Burnquist, M. E. Sorrells, J. E. Irvine and A. H. Paterson. 1998. Detailed alignment of *Saccharum* and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Mudge, J., W. R. Andersen, R. L. Kehrer and D. J. Fairbanks. 1996. A RAPD genetic map of *Saccharum officinarum*. Crop Sci. 36:1362-1366

Reffay, N., P. A. Jackson, K. S. Aitken, J. Y. Hoarau, A. D'Hont, P. Besse and C. L. McIntyre. 2005. Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. Mol. Breed. 15: 367–381.

Sreenivasan, T. V., B. S. Ahloowalia and D. J. Heinz. 1987. Cytogenetics in: Heinz D. J. (ed), Sugarcane improvement through breeding. Elsevier, Amsterdam. pp. 211-253

CHAPTER 2 TARGET REGION AMPLIFICATION POLYMORPHISM (TRAP) FOR ASSESSING GENETIC DIVERSITY IN SUGARCANE GERMPLASM COLLECTIONS

2.1 Introduction

The genus *Saccharum* is composed of six species, namely *S. officinarum* L., *S. barberi* Jesw., *S. sinense* Roxb., *S. spontaneum* L., *S. robustum* Brandes and Jeswiet ex Grassl, and *S. edule* Hassk. (Brandes, 1958). The modern *Saccharum spp* (cultivated sugarcane) is believed to have originated from complex hybridization events (termed 'nobilization') between *Saccharum officinarum*, *S. barberi, S. sinense* and the wild related species *S. spontaneum* (Sreenivasan et al., 1987). Until the end of the 19th century, cultivated sugarcane comprised mainly of the vegetatively propagated *S. officinarum* (the main sugar producing cane) together with *S. barberi* and *S. sinense* (Jannoo et al., 1999). *Saccharum officinarum*, however, is believed to have evolved through hybridization of species such as *Erianthus arundinaceus*, *S. spontaneum* and *S. robustum* (Daniels *et al*., 1975) whereas *S. barberi* and *S. sinense* are believed to be natural hybrids between *S. officinarum* and *S. spontaneum* (Daniels and Roach, 1987). Mukherjee (1957) coined the term *Saccharum* complex to encompass four closely related interbreeding genera *viz*., *Saccharum, Erianthus* (=sect. Ripidium), *Narenga* and *Sclerostachya,* all of which are supposedly implicated in the origin of sugarcane. Daniels et al. (1975) revised this grouping to include *Miscanthus* sect. Diandra Keng but the phylogenetic relationship between members of the group remains unclear (Irvine, 1999).

A better understanding of the genetic diversity and inter-relationships among members of the *Saccharum* complex will facilitate exploitation of this germplasm in improving sugarcane. Traditional methods which combined agronomic and morphological characteristics have been useful in identifying and describing differences

between members of the *Saccharum* complex (Artschwager and Brandes, 1958; Skinner, 1972; Skinner et al. 1987). However, members of the *Saccharum* complex are predominantly outcrossing and are maintained by vegetative propagation. As such, they are highly heterozygous and display enormous plasticity in the phenotypic expression of traits. Although morphological traits can be used to identify and classify clones, most of the traits are influenced by the environment under which the clones are grown or selected. Variability caused by genotype x environment interactions and inadvertent mislabeling of clones can adversely influence data derived from phenotypic evaluation and clonal records.

With the advent of molecular markers it is now possible to make direct inferences about genetic diversity and inter-relationships among organisms at the DNA level without the confounding effects of the environment and/or faulty pedigree records. Indeed, a vast number of molecular marker techniques such as isoenzymes (Glaszmann et al., 1989), RFLP (D'Hont et al., 1994; Jannoo et al., 1999; Coto et al., 2002), ribosomal DNA (Glaszmann et al., 1990 ; Pan et al., 2000), microsatellites (Piperidis et al., 2001; Cordeiro et al., 2003), AFLP (Besse et al., 1998; Lima et al., 2002) and molecular cytogenetics (D'Hont et al., 1996) have been instrumental in explaining genetic diversity and inter relationships among accessions in sugarcane germplasm collections.

The underlying goal for studying genetic diversity and inter-relationships among germplasm collections is to eventually use that information to facilitate the development of better performing varieties of the cultivated species. The results from genetic diversity studies may, therefore, be more useful if the segments of the genome sampled or measured correspond to segments bearing genes of interest to the breeding program. Current molecular marker tools, such as RFLP, RAPD, AFLP, gSSR (SSRs derived from

genomic as opposed to EST sequences), have unarguably been very useful in dissecting the level and pattern of genetic diversity in sugarcane germplasm collections. However, the polymorphism generated by these marker techniques are randomly distributed across the genome and only those that can be associated with traits through QTL studies would be of immediate interest to the breeder. Even when QTL analysis is performed, the underlying association is often based on relatively large linkage blocks. Transferability of QTLs between populations remains a looming question in the minds of many plant breeders.

Sugarcane remains a complex and recalcitrant crop to study and improve using genetics approaches owing to the large genome size, high ploidy level and complicated genome organization. However, recent access to increasing numbers of sugarcane EST sequences obtained from diverse cDNA libraries coupled with freely available bioinformatics tools now allow us to explore new opportunities in sugarcane molecular marker research. The Target Region Amplification Polymorphism (TRAP) is a simple PCR-based marker technique which uses EST or gene information to generate polymorphism (Hu and Vick, 2003). A fixed primer of about 18 nucleotides is designed from EST sequences or genes of interest and an arbitrary primer of about the same length is designed with either an AT- or GC-rich motif to anneal with an intron or exon, respectively (Li and Quiros, 2001; Hu and Vick, 2003). TRAP markers have not previously been used to genotype sugarcane. Therefore, the objective of this study was to evaluate the potential of TRAP markers for assessing genetic diversity in sugarcane germplasm collections.

2.2 Materials and Methods

2.2.1 Plant Material and DNA Extraction

Thirty genotypes, representing three genera namely *Saccharum*, *Miscanthus* and *Erianthus* were used in the study (Table 2.1). Representing *Saccharum* species were S. *officinarum, S. barberi*, *S. sinense, S. spontaneum, S. robustum* as well as cultivars, cultivar-derived mutants and interspecific hybrids. The genotypes DW1 and DW2 are cultivar-derived genetic mutants from the cultivar LCP81-137 (Burner, 1999). The genotypes 16 Low and 40 High are first generation interspecific hybrids from a cross between LA Stripe (S. *officinarum*) x SES 147b (*S. spontaneum*) and are being retained in the collection because of their low and high sucrose content, respectively. These genotypes form part of the germplasm collection maintained at the USDA Sugarcane Research Unit at Houma, Louisiana.

Young leaves were collected from each genotype, frozen immediately in liquid nitrogen and stored at -80 C. The leaves were later ground to a powder in liquid nitrogen. Genomic DNA was extracted using the Plant DNeasy Mini Kit (Qiagen, Valencia, CA) following the manufacturer's protocol. Concentrations of extracted DNA were estimated by known concentration of lambda DNA in 1% agarose gel.

2.2.2 TRAP Markers

2.2.2.1 Primer Design

The design of fixed primers was based on the method reported by Hu and Vick (2003). The nucleotide sequences of six genes of interest were obtained from the GenBank database at NCBI. Of the six selected genes, five are believed to be involved in carbohydrate (sucrose) metabolism while the remaining one is believed to play an important role in cold tolerance. The primers were designed using the Primer3 software

(http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Rozen and Skaletsky, 2000), out of which only the forward primer was used as a fixed primer. The primer optimum

Table 2. 1. Description of the 30 genotypes used for TRAP marker analysis

Number	Clone name	Genera or species †	Code
1	Kalingpong	*Erianthus*	Er
2	Dwarf 1	*Saccharum* species hybrid (Cultivar-derived mutant)	DW1
3	Dwarf 2	*Saccharum* species hybrid (Cultivar-derived mutant)	DW2
4	16 Low	*Saccharum* species hybrid (F_1 between *S. officinarum and S. spontaneum*)	Hy1
5	40 High	*Saccharum* species hybrid (F_1 between *S. officinarum and S. spontaneum*)	Hy2
6	POJ2878	*Saccharum* species hybrid (Cultivar)	Cu1
7	LCP 85-384	*Saccharum* species hybrid (Cultivar)	Cu2
8	CP 77-310	*Saccharum* species hybrid (Cultivar)	Cu3
9	CP 77-407	*Saccharum* species hybrid (Cultivar)	Cu4
10	LCP 85-845	*Saccharum* species hybrid (Cultivar)	Cu5
11	Miscanthus	*Miscanthus*	Mi
12	Ganapathy	*S. barberi*	Sb1
13	Chin	*S. barberi*	Sb2
14	La Stripe	*S. officinarum*	So1
15	La Purple	*S. officinarum*	So2
16	Cuba	*S. officinarum*	So3
17	IN84-068A	*S. officinarum*	So4
18	NG 57-54	*S. robustum*	Sr1
19	NG 57-159	*S. robustum*	Sr2
20	Molokai 5573	*S. robustum*	Sr3
21	IMP72-232	*S. robustum*	Sr4
22	NG77-218	*S. robustum*	Sr5
23	Chukche	*S. sinense*	Ssi
24	SES 147b	*S. spontaneum*	Ssp1
25	Coimbatore	*S. spontaneum*	Ssp2
26	MPTH97-213	*S. spontaneum*	Ssp3
27	MPTH97-200	*S. spontaneum*	Ssp4
28	MPTH97-107	*S. spontaneum*	Ssp5
29	PIN84-1B	*S. spontaneum*	Ssp6
30	Molokai1032B	*S. spontaneum*	Ssp7

† The original sugarcane cultivars were derived from crossing mainly between *S. officinarum and S. spontaneum* followed by several generations of backcrosses to *S. officinarum*. Present day cultivars are selections from cultivar x cultivar crosses.

size, maximum size and minimum size were set to 18 nt. The optimum T_m, maximum T_m and minimum T_m were set to 53°C, 55°C and 50°C respectively. The genes, GenBank accession numbers and designed primer sequences used in this study are given in Table 2. 2.

Table 2. 2. Sequences of fixed and arbitrary primers used for TRAP markers.

	Gene	Sequence (5'→ 3')	GenBank accession number
Fixed primers	Sucrose Synthase (SuSy)	GGAGGAGCTGAGTGTTTC	AF263384
	Sucrose Phosphate Synthase (SuPS)	CGACAACTGGATCAACAG	AB001338
	Pyruvate Orthophosphate DiKinase (PODK)	CGTAAAGATTGCTGTGGA	AF194026
	Soluble Acid Invertase (SAI)	AGGACGAGACCACACTCT	AF062735
	Calcium Dependent Protein Kinase (CDPK)	ACAGAACCACCAAAGGAG	CF572977
	Starch Synthase (StSy)	GGCAAGAAGAAGTTCGAG	AF446084
Arbitrary primers	Arbi 1	GACTGCGTACGAATTAAT	
	Arbi 2	GACTGCGTACGAATTGAC	
	Arbi 3	GACTGCGTACGAATTTGA	

Arbitrary reverse primer sequences were obtained from Li and Quiros (2001). These primers comprise three selective nucleotides at the 3' end, 4 nucleotides of AT- or GC-rich content in the core region and 11 nucleotides as filler sequences at the 5' end. In addition, the basic rules of primer design such as self-complementarity and maintenance of 40-60% GC content were upheld (Table 2. 2). The AT- and GC-rich primers supposedly target introns and exons, respectively (Li and Quiros, 2001).

2.2.2.2 PCR Protocol

TRAP reactions were performed based on the protocol of Hu and Vick (2003). Fixed primers were combined with each of three arbitrary primers for a total of 18 primer combinations. Each reaction was carried out in a total volume of 20μL containing 2μL of 10x PCR buffer, 1.0μL of 25mM $MgCl_2$, 1.0μL each of 10μM fixed and arbitrary primers, 1.0μL of 10mM dNTPs (Promega, Madison, WI), 0.35μL of 5U *Taq* polymerase (Promega, Madison, WI) and 1.0μL of 50-80ng genomic DNA. The conditions for PCR were as follows: an initial denaturing step was performed at 94^{o}C for 4 min followed by 5 cycles at 94^{o}C for 45 s, 35^{o}C for 45 s and 72^{o}C for 1 min, followed by 35 cycles at 94^{o}C for 45 s, 53^{o}C for 45 s and 72^{o}C for 1 min with a final extension step at 72^{o}C for 7 min. All the PCR reactions were performed on an *i-cycler* (BioRad Labs, Hercules, CA). After PCR, the amplified products were run on 7% polyacrylamide denaturing gel for 2.0 hrs at 110 W. Silver staining procedure was employed to develop the gel and to detect the bands.

2.2.3 Statistical Analysis

Bands from the TRAP gel were scored, as '1' for presence and '0' for absence, in all genotypes. Only readable bands were scored while ambiguous bands were ignored and excluded from the analysis. Allelic diversity at a given locus can be measured by Polymorphism Information Content (PIC) wherein a marker can distinguish two alleles taken at random from a population and it was calculated as follows:

$$PIC = 1- \Sigma f_i^2$$

where, f_i is the frequency of the i^{th} allele (Weir 1990). Considering the number of alleles at a locus along with their relative frequencies in a given population, an estimate of the

discriminatory power of a marker can be obtained by calculating the PIC (Vuylsteke et al 2000). Jaccard-similarity coefficient (1908) was used to calculate the estimate of genetic similarity (GS) among pairs of genotypes as follows: $GS_{ij} = a/(a+b+c)$ where GS_{ij} is the genetic similarity measurement between individuals *i* and *j*, the number of matching bands in both individuals is represented by *a* whereas *b* and *c* are the number of bands present in individual *i* and *j* respectively but not in their counterparts. The GS matrix was used to perform cluster analysis using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) algorithm (Sneath and Sokal, 1973) following the Sequential Agglomerative Hierarchical Nested (SAHN) cluster analysis. The co-phenetic values were calculated to test the goodness of fit between the clusters in the dendrograms and the similarity index matrix. In addition, Principal Coordinate Analysis (PCoA) was performed to supplement the findings obtained from cluster analysis. All the above analyses were performed employing different modules of NTSYS-PC software, version 2.11L (Rohlf, 2000). For the purpose of comparison between clusters and also to determine the robustness of the cluster, bootstrap analysis was done with 10,000 replications using the PAUP version 4.0b10 software (Sinauer Associates Inc., MA) which employs Nei and Li (1979) method for cluster development.

2.3 Results and Discussion

2.3.1 TRAP Marker Polymorphism and PIC Values

The summary of TRAP markers produced by the 18 primer combinations (six fixed forward primers in combination with three arbitrary reverse primers) across all 30 genotypes is given in Table 2. 3. The 18 primer combinations generated a total of 600 unambiguous bands of which 529 (88%) were polymorphic. The bands ranged in size from 100 to 700 bp. The number of bands detected by individual primer combinations

ranged from 15 (SuSy + Arbi1) to 58 (CDPK + Arbi 3) with an average of 33. These two primer combinations were also responsible for the least (14 in SuSy + Arbi1) and the most (48 in CDPK + Arbi 3) number of polymorphic bands produced for an average of 29 polymorphic bands per primer combination. Polymorphism was generally high (>

Table 2. 3. Total number of bands, number of polymorphic bands, percent polymorphism and Polymorphism Information Content (PIC) for each of 18 TRAP primer combinations.

Primer combination	Bands observed	Polymorphic bands	Percent Polymorphism	PIC value
SuSy + Arbi 1	20	20	100.00	0.32
SuSy + Arbi 2	32	32	100.00	0.28
SuSy + Arbi 3	15	14	93.33	0.20
SuPS + Arbi 1	19	17	89.47	0.33
SuPS + Arbi 2	48	47	97.91	0.24
SuPS + Arbi 3	29	21	72.42	0.14
SAI + Arbi 1	39	34	94.87	0.26
SAI + Arbi 2	28	28	100.00	0.22
SAI + Arbi 3	46	40	86.95	0.21
StSy + Arbi 1	41	31	75.60	0.20
StSy + Arbi 2	50	41	82.00	0.24
StSy + Arbi 3	28	28	100.00	0.21
PODK + Arbi 1	36	31	86.11	0.27
PODK + Arbi 2	29	22	75.86	0.36
PODK + Arbi 3	32	27	84.37	0.29
CDPK + Arbi 1	29	29	100.00	0.11
CDPK + Arbi 2	21	19	90.47	0.25
CDPK + Arbi 3	58	48	82.75	0.23
Total	600	529		
Average	33.33	29.38	88.14	0.24

50%), ranging from 72 to 100%. The high level polymorphism could be attributed to the complex genetic structure of sugarcane with high levels of polyploidy and

heterozygosity. Similar high levels of polymorphism have been reported in *Saccharum* species by Besse et al (1998) and Lima et al (2002) using AFLP markers. The Polymorphism Information Content (PIC), which measures information content as a function of a marker system's ability to distinguish between genotypes (Weir, 1990), varied among the primer combinations ranging from 0.11 in CDPK + Arbi 1 to 0.36 in PODK + Arbi2 with an average of 0.24. The PIC values indicate a good discriminatory power of the dominant TRAP marker system. Comparable PIC values have been reported using dominant markers like RAPD and AFLP in African plantain (Ude et al., 2003) and AFLP in wheat (Bohn et al., 1999).

For the 18 TRAP primer combinations, the Jaccard's GS estimates ranged from 0.33 (Kalingpong and *Miscanthus*) to 0.94 (Dwarf 1 and Dwarf 2) with a mean of 0.68. A dendrogram with a co-phenetic value of 0.96 was generated (Fig. 2.1) based on 435 pair-wise GS estimates. A co-phenetic value of > 0.80 is said to indicate a strong goodness of fit for dendrograms (Rohlf, 2000). Bootstrap, based on 10,000 re-samplings of the data set, and cluster analyses following the Nei and Li (1979) method produced a similar dendrogram (data not shown). This further confirmed the robustness of the dendrogram obtained by the UPGMA method based on Jaccard's similarity coefficient (Fig 2.1).

2.3.2 Genetic Diversity and Relationships among Genotypes

Genetic diversity and relationships among the genotypes in this study were depicted by both cluster and PCoA (Figs. 2.1 and 2. 2). A separate analysis was performed for the three cold tolerance-related primer combinations (i.e. CDPK/ Arbi1, 2 and 3). As the results did not differ from the one derived from using the 15 sucrose-related primer combinations, the data were merged and used for one combined analysis. The dendrogram from cluster analysis revealed two distinct groups among the *Saccharum*

species. Group I comprised the genotypes representing *S. officinarum*, *S. sinense, S. barberi, S. robustum* along with cultivars, cultivar-derived mutants and hybrids while Group II comprised all the *S. spontaneum* genotypes. The single *Erianthus*

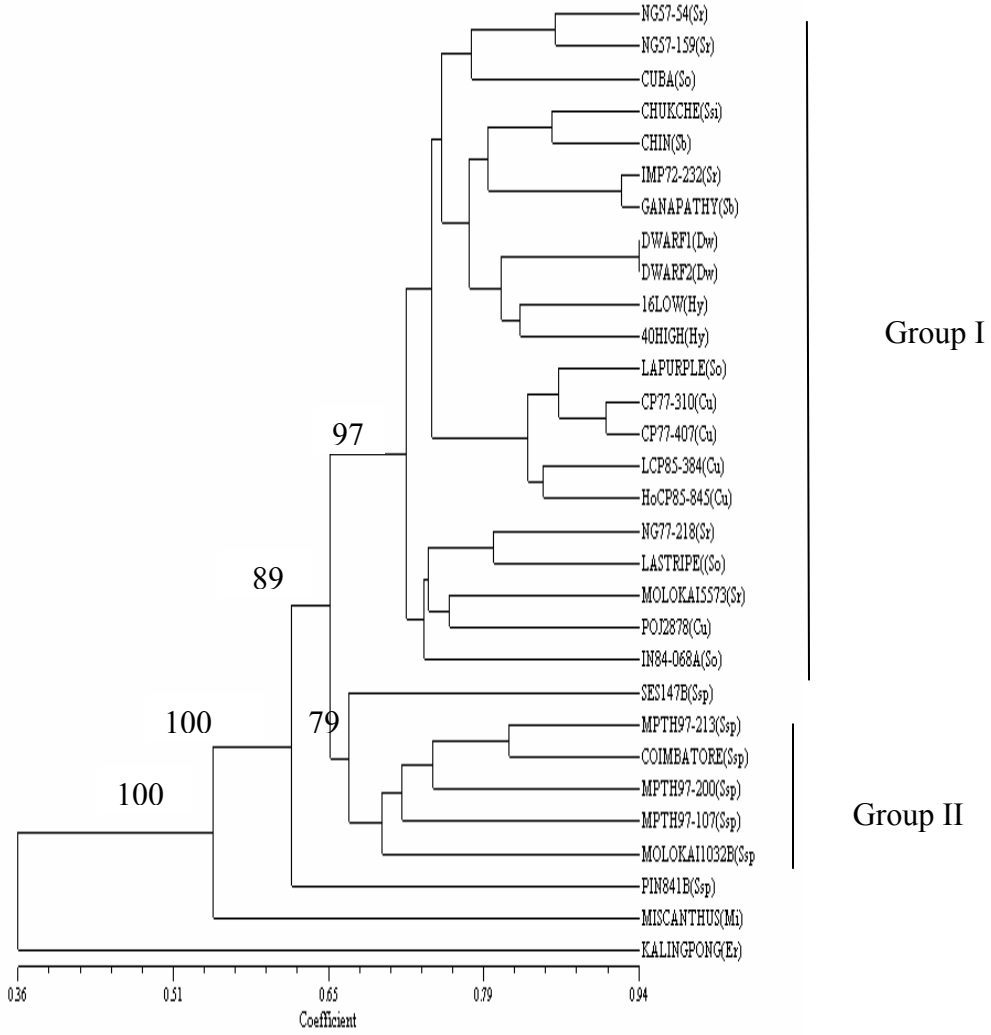


Figure 2.1. Grouping among 30 genotypes from a sugarcane germplasm collection based on 18 TRAP primer combinations using the UPGMA method. Numbers represent values from boostrap analysis. Abbreviations: Mi, *Miscanthus*; Er, *Erianthus*; Cu, Cultivar; Ssp, *Saccharum spontaneum*; So, *S. officinarum;* Sr, *S. robustum*; Sb, *S. barberi*; Ssi, *S. sinense*; Hy, Hybrid; Dw, Dwarf.

(Kalingpong) and *Miscanthus* genotypes were each placed distinctly in the dendrogram supporting the taxonomic evidence which assigned each of them to a separate genus

(Daniels et al., 1975). Remarkably similar results were obtained from the PCoA (Fig. 2. 2). The first three axes in the PCoA explained a cumulative variation of 42.23 percent. As with the cluster analysis, all the *S. spontaneum* genotypes formed a well individualized group while the rest of the *Saccharum* species along with the cultivars, cultivar-derived mutants and hybrids clustered together as one interrelated group. *Miscanthus* and *Erianthus* (Kalingpong) were placed distinctly, again lending credence to the claim that they are separate genera.

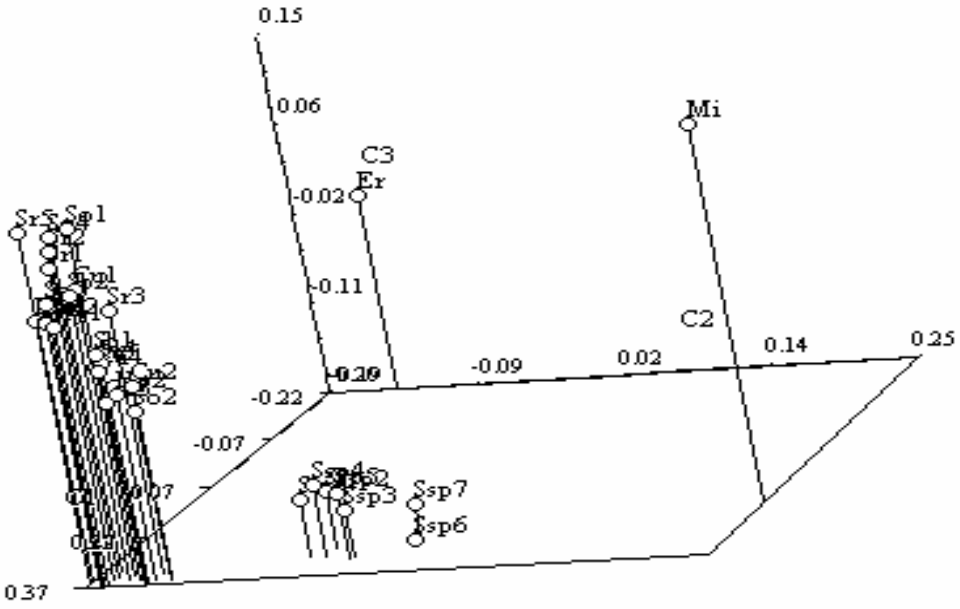


Figure 2. 2. Association among 30 genotypes from a sugarcane germplasm collection as revealed by PCoA of genetic distances based on 18 TRAP primer combinations. Abbreviations: Mi, *Miscanthus*; Er, *Erianthus*; Cu, Cultivar; Ssp, *Saccharum spontaneum*; So, *S. officinarum;* Sr, *S. robustum*; Sb, *S. barberi*; Ssi, *S. sinense*; Hy, Hybrid; Dw, Dwarf. Numbers were used to uniquely identify a genotype (for example Ssp6) when there was more than one genotype representing a species or group.

The strong differentiation between *Erianthus* and *Saccharum* genera as revealed by TRAP markers was previously demonstrated using rDNA spacers (Al-Janabi et al., 1994), RFLP (Burnquist et al., 1992), AFLP (Besse et al., 1998), 5S RNA intergenic spacers (Pan et al., 2000), and sugarcane- (Cordeiro et al., 2003) as well as maize- (Selvi et al., 2003) derived microsatelite markers. Similarly, evidence from microsatellite markers (Cordeiro et al., 2003) and 5S RNA intergenic spacers (Pan et al., 2000) had previously been used to document the distinction between *Miscanthus* and *Saccharum* species.

Although no distinct sub-groups were found within Group I, the clustering of genotypes in this group seems to be in accordance with the ancestral relationships among these species (Fig 2. 1). *Saccharum robustum* is believed to be one of the progenitors of *S. officinarum* (Brandes, 1958; Daniels and Roach, 1987). Significant similarities have been reported between *S. robustum* and *S. officinarum* with regards to morphology, cytology and physiology, however they differed in fiber and sugar content (Irvine, 1999). The high degree of similarity between *S. robustum* and *S. officinarum* has also been revealed using RAPD (Nair et al., 1999) and microsatellite markers derived from maize (Selvi et al., 2003).

Saccharum barberi and *S. sinense* genotypes were found within the same sub-group albeit along with a *S. robustum* genotype. *Saccharum barberi* and *S. sinense* might differ enough to be distinct (Glaszmann et al., 1990) but Whalen (1991) contends that those minor differences are not enough to classify them as two separate species. Moreover, these two species are thought to be interspecific hybrids between *S. officinarum* and *S. spontaneum* (Brandes, 1958; Daniels and Roach, 1987) and this has been substantiated using evidence from RFLP (Lu et al., 1994), RAPD (Nair et al., 1999),

maize-derived microsatellite markers (Selvi et al., 2003) and GISH analyses (D'Hont, 1993).

It was also not surprising that cultivars were found in Group I, indicating their closer relationship with *S. officinarum* compared to *S. spontaneum*. Most of the cultivars bred after the turn of the 20th century are interspecific hybrids between *S. officinarum* and *S. spontaneum*. However, cultivars inherited a greater proportion of the *S. officinarum* genome as 'nobilization' involved several backcrosses to the *S. officinarum* parent during which this parent transmitted the somatic chromosome number to its progeny (Bhat and Gill, 1985; Bremer, 1961; Sreenivasan et al., 1987; D'Hont et al., 1996).

The closest relationship in the dendrogram was found between the two cultivar-derived dwarf mutants which, is in agreement with the origin of these genotypes. Except for the legendary cultivar POJ2878, all the contemporary cultivars were found in the same sub-group albeit with a *S. officinarum* genotype. This is hardly surprising since these contemporary cultivars are more closely related relative to POJ2878. However, it was interesting to note that within this sub-group, LCP85-384 clustered closer to HoCP85-845 than it did to either of its parents, namely CP77-310 (female) and CP77-407 (male). LCP85-384 and HoCP85-845 share a common heritage in that their grand parents are full siblings. But the closer association between these genotypes, relative to that between LCP85-384 and its parents, is possibly due to the effects of breeding and selection which is not accounted for by pedigree history. Furthermore, the primers employed in this study were designed to preferentially amplify a small segment of the genome, that is, segments associated with sucrose content and cold tolerance. The effect of selection, especially for sucrose related genes, coupled with the fact that only a small

portion of the genome was being assayed could perhaps explain why the genotypes in Group I failed to form distinct subgroups and clustered instead as one interrelated group. In a study using maize-derived microsatellite markers, *S. barberi* and *S. sinense* genotypes grouped together but the group was placed in between the *S. officinarum* and *S spontaneum* clusters (Selvi et al., 2003). The authors used this as evidence to suggest that both *S. officinarum* and *S spontaneum* were involved in the ancestry of these two species.

Table 2. 4. Mean genetic similarity (GS) within and between *Saccharum* species

	S. officinarum	*S. spontaneum*	*S. robustum*	Cultivars
S. officinarum	0.71†			
S. spontaneum	0.62	0.68		
S. robustum	0.74	0.63	0.76	
Cultivars	0.74	0.66	0.73	0.80

† The numbers in the diagonal are for 'within' GS estimates

The average GS within and among groups of genotypes or species was computed as an additional measure to assess genetic diversity (Table 2. 4). Only groups or species represented by at least four genotypes were considered. The estimates showed that the least amount of similarity existed among the *S. spontaneum* genotypes (0.68) indicating the relatively higher level of heterozygosity and polymorphism that exist within this species. *Saccharum spontaneum* is generally accepted as the most diverse of the *Saccharum* species in terms of geographical distribution, chromosome number (2n = 40-128) and morphology (Daniels and Roach, 1987). *Saccharum spontaneum* is considered an untapped resource for sugarcane germplasm improvement in Louisiana. The major focus has been on traits such as disease resistance, cold tolerance and ratooning ability although recent evidence using molecular markers suggest that wild relatives such as *S. spontaneum* (with relatively low sucrose content) cannot be discounted as potential

contributors of novel genes for traits such as sucrose content (Tanksley and McCouch, 1996; Reffay, 2005). TRAP markers could potentially be useful for identifying novel variation and for introgression breeding.

Compared to *S. spontaneum*, genetic similarity was higher among *S. officinarum* < *S. robustum* < cultivars. Genetic diversity has generally been reported as being very low among cultivated sugarcane as very few progenitor clones were involved in the initial 'nobilization' event and the products from 'nobilization' became the foundation clones for most breeding programs.

The highest genetic similarity among groups was obtained between cultivars and *S. officinarum* (0.74) followed by *S. robustum* and *S. officinarum* (0.74). These results provide additional support that, *S. robustum* is a likely progenitor of *S. officinarum* (Sreenivasan et al., 1987) and that cultivars inherited most of their sucrose related genes from *S. officinarum.* The least amount of similarity was observed between *S. officinarum* and *S. spontaneum* (0.62) reflecting the distinctness of these two species.

2.3.3 Genus and/or Species Specific Markers

Generally, very few bands were discrete across species or genus. The main types of uniqueness found were situations where a band was either present or absent among all genotypes of a species, but, the same band was polymorphic among the other species or genotypes. For example, whereas a Susy + Arbi 2 (500-600bp) fragment was polymorphic among *S. spontaneum* and cultivars, this fragment was uniquely absent in all the *S. officinarum, S. robustum, S. barberi*, *S. sinense*, *S. officinarum* x *S. spontaneum* hybrids, cultivar-derived dwarf genotypes and present in the two *Erianthus* and *Miscanthus* genotypes. Fragment Susy + Arbi 3 (350-400bp) was absent in all *S. spontaneum, Erianthus* and *Miscanthus* genotypes and present among the rest of the

genotypes except among cultivars where it was present in three of the five genotypes. But the most significant fragment was SuPS + Arbi1 (600-700bp) which was present in the two *Erianthus* and *Miscanthus* genotypes but more importantly in all the cultivars and *S. officinarum* genotypes and absent in all *S. spontaneum* genotypes. Fragments which are unique to either *S. officinarum* or *S. spontaneum* are of significance to the Louisiana introgression breeding program because *S. officinarum* and *S. spontaneum* are being used as sources of genes to increase sucrose content and stress tolerance. Another interesting fragment was produced by SAI + Arbi 2 (600-700bp) which was absent among all the *Saccharum* species and present in the two *Erianthus* and *Miscanthus* genotypes. Such a fragment could be useful in distinguishing *Saccharum* from other genera.

2.3.4 Sequencing of Amplified Products

TRAP amplicons were sequenced in an effort to verify if indeed the TRAP marker binding sites correspond to candidate genes. The bands were excised from a PAGE gel, re-amplified and sequenced directly. We sequenced bands from SES 147b a *S. spontaneum* genotype. A fragment of 535 bp from *S. spontaneum,* amplified by the StSy + Arbi3 primer combination, showed homology (E = 5.7) with an EST (Accession # AF079258) of a *Sorghum bicolor* granule-bound starch synthase gene. A similar level of homology (E = 5.7) was found with the ESTs of the granule-bound starch synthase genes of *Cymbopogon pospischilii* (Accession # AF079248)*, Heteropogon contortus* (Accession # AF079253) and *Coelorachis selloana* (Accession # AY062271.1). Much higher levels of homology were found with the mRNA sequence of *Zea mays* endosperm transcriptome (E = 6e-37; Accession # BT018673.1) and a cDNA clone corresponding to chromosome 3 of *Oryza sativa* (E= 2e-15; Accession # AK105342.1).

We undertook further analyses of the sorghum EST sequence as sorghum is considered a relative of sugarcane based on comparative mapping studies (Paterson et al., 1995; Ming et al., 1998; Ming et al. 2002). The sorghum EST sequence showed a 100 % match with a segment of the 535 bp sequence from *S. spontaneum*. A BLASTn search using the sorghum sequence pulled up, among other sequences, the same EST of *Saccharum officinarum* (Accession # AF446084) from which the 'StSy' fixed primer was designed.

Another fragment of 295 bp from *S. spontaneum,* amplified by the CDPK + Arbi2 primer combination showed homology (E = 0.19) with a segment of mRNA corresponding to a putative receptor-like protein kinase gene of *Oryza sativa.* In addition it showed homology (E = 0.19) with a segment of a clone from chromosome 5 of *Oryza sativa* containing a putative receptor-like protein kinase gene. The need is obvious for further sequencing analyses, using more than one genotype to demonstrate that similar-sized TRAP bands are allelic.

2.4 Conclusions

Our results provide support for the utility of TRAP markers for assessing genetic diversity in sugarcane germplasm collections. TRAP primers are designed from ESTs or gene sequences; thus, the potential to generate polymorphism around targeted gene sequences is an attractive feature of TRAP markers. Although the TRAP markers reported here have yet to be mapped in sugarcane, results from sequencing and BLASTn analyses of TRAP amplicons lend some support to the proposal that TRAP primers may be targeting gene regions. This was further substantiated by the aggregation of genotypes in the study which seemed to reflect the fact that the TRAP primers were designed from sucrose- and cold tolerance-related gene/EST sequences. Genotypes of the relatively low

sucrose and cold tolerant species, *S. spontaneum,* formed one distinct group. Whereas, in contrast to previous studies, genotypes of the high sucrose and cold sensitive *Saccharum* species including S. *officinarum, S. barberi*, *S. sinense,* and *S. robustum,* formed one inter-related cluster with no distinct sub-groups. TRAP markers could potentially be useful in the characterization and management of domesticated and wild germplasm where the aim is to enhance the germplasm for specific traits. Genetic diversity could be evaluated using TRAP markers for the trait(s) of interest and genotypes or species displaying unique diversity selected for germplasm enhancement. We are currently employing TRAP markers in a QTL mapping study in an effort to further authenticate their potential to target gene regions.

2.5 References

Al-Janabi, S. M., B. M. S. Sobral, C. Peterson, and M. McClelland. 1994. Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: Saccharinae. Theor. Appl. Genet. 88: 933-944.

Artschwager, E and E. W. Brandes. 1958. Sugarcane (*S. officinarum* L.). Origin, classification, characteristics and descriptions of representative clones USDA Agri. Handbook. 122 pp 307.

Besse, P., G. Taylor, B. Carroll, N. Berding, D. Burner, and C. L. McIntyre.1998. Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. Genetica 104: 143-153

Bhat, S. R, and B. S. Gill. 1985. The implication of 2n egg gametes in nobilisation and breeding of sugarcane. Euphytica 34:377-384.

Bohn, M., H. F. Utz and A. E. Melchinger.1999. Genetic similarities among wheat cultivars determined on the basis of RFLPs, AFLPs and SSRs and their use for predicting progeny variance. Crop Sci. 39: 228-237

Brandes, E. W. 1958. Origin, classification and characteristics. In: E Artschwager and EW Brandes, Sugarcane (*S. officinarum* L.). USDA Agri. Handbook. 122 pp 1-35, 260-262

Bremer, G. 1961. Problems in breeding and cytology of sugarcane. Euphytica 10: 59-78.

Burner, D. M. 1999. Registration of Dwarf1 Sugarcane Genetic stock. Crop Sci. 39: 309

Burnquist, W. L., M. E. Sorells and S. D. Tanksley.1992. Characterization of genetic variability in sacccharum germplasm by means of restriction fragment length polymorphism (RFLP) analysis. XXI Proc Int Soc. Sugar Cane Technol. 21: 355-365

Cordeiro, G. M., Y. B. Pan and R. J. Henry. 2003. Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. Plant Sci. 165: 181-189

Coto, O., M. T. Cornide, D. Calvo, E. Canales, A. D'Hont and F. de Prada. 2002. Genetic diversity among wild sugarcane germplasm from Laos revealed with markers. Euphytica 132: 121-130.

D'Hont, A., Y. H. Lu, D. Gonzales-de-Leon, P. Feldman, C. Lanaud and J. C. Glaszmann. 1994. A molecular aproach to unravelling the genetics of sugarcane, a complex polyploid of the Andropogonae tribe. Genome 37: 222-230.

D'Hont, A., L. Grivet, P. Feldman, P. S. Rao, N. Berding and J. C. Glaszmann. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum spp*.) by molecular cytogenetics. Mol. and Gen Genet. 250:405-413

D'Hont, A., Y. H. Lu, P. Feldmann and J. C. Glaszmann. 1993. Cytoplasmic diversity in sugarcane revealed by heterologous probes. Sugar Cane 1: 12-15.

Daniels, J and B. T. Roach. 1987. Taxonomy and evolution in sugarcane, in: H. DJ (ed), Sugarcane improvement through breeding. Elsevier, Amsterdam. pp. 7-84

Daniels, J., P. Smith, N. Paton and C. A. Williams. 1975. The origin of the genus Saccharum. Sugarcane Breed. Newsl. 36: 24-39

Glaszmann, J. C., A. Fauret, J. L. Noyer, P. Feldman and C. Lanaud. 1989. Biochemical genetic markers in sugarcane. Theor. Appl Genet. 78: 537-543

Glaszmann, J. C., Y. H. Lu and C. Lanaud. 1990. Variation of nuclear ribosomal DNA in sugarcane. J. Genet. Breed. 44: 191-198

Hu, J.G and B. A. Vick. 2003. Target Region Amplification Polymorphism: A novel marker technique for plant genotyping. Plant Mol. Biol. Reporter 21: 289-294.

Irvine, J. E.1999. *Saccharum* species as horticultural classes. Theor. Appl. Genet. 98: 186-184

Jaccard, P. 1908 Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44 : 223-270

Jannoo, N., L. Grivet, M. Seguin, F. Paulet, R. Domaingue, P. S. Rao, A. Dookun A, D'Hont, A and J. C. Glaszmann. 1999. Molecular investigation of the genetic base of sugarcane cultivars. Theor. Appl. Genet. 99: 171-184

Li, G and C. F. Quiros. 2001. Sequence related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theor. Appl. Genet. 103: 455-461

Lima, M. L. A., A. A. F. Garcia, K. M. Oliveira, S. Matsuoka, H. Arizono, C. L. de Souza Jr and A. P. de Souza. 2002. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum spp*). Theor. Appl. Genet. 104: 30-38

Lu, Y. H., A. D'Hont, D. I. T. Walker, P. S. Rao, P. Feldman and J. C. Glaszmann. 1994. Relationships among ancestral species of sugarcane revealed by RFLP using single –copy maize nuclear probes Euphytica 78: 7-18

Ming, R., S. C. Liu, J. E. Bowers, P. H. Moore, J. E. Irvine and A. H. Patterson. 2002. Construction of a *Saccharum* consensus genetic map from two interspecific crosses. Crop Sci. 42:570-583

Ming, R., S. C. Liu, Y. R. Lin, D. Braga, J. daSilva, A. van Deynze, T. F. Wenslaff, K. K. Wu, P. H. Moore, W. Bunquist, M. E. Sorrells, J. C. Irvine and A. H. Patterson. 1998. Alignment of *Sorghum* and *Sacharrum* chromosomes : Comparative organization of closely related diploid and polyploidy genomes. Genetics 150:1663-1882.

Mukherjee, S. K.1957. Origin and distribution of *Saccharum*. Bot. Gaz 119: 55-61

Nair, N. V., S. Nair, T. V. Sreenivasan and M. Mohan.1999. Analysis of genetic diversity and phylogeny in Saccharum and related genera using RAPD markers. Genet. Resources Crop Evol. 46: 73-79

Nei, M and W. H. Li.1979. Mathematical Model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci., USA 96: 5269-5273

Pan, Y. B., D. M. Burner and B. L. Legendre. 2000. An assessment of the phylogenetic relationship among sugarcane and related taxa based on the nucleotide sequence of 5S rRNA intergenic spacers. Genetics 108: 285-295

Paterson, A. H., Y. R. Lin, Z. K. Li, K. F. Schertz, J. F. Dobley , S. R. M. Pinson, S. C. Liu, J. W. Stansel and J. E. Irvine.1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714-1718.

Piperidis, G., G. O. Taylor and G. R. Smith. 2001. A Microsatellite marker database for fingerprinting sugarcane clones. In: XXIV Proc. Intl. Soc. Sugarcane Technol. Pp 632-633.

Reffay, N., P. Jackson, K. Aitken, J Hoarau, A. D'Hont, P. Besse and C. L. McIntyre. 2005. Characterization of genome regions incorporated from an important wild relative into Australian sugarcane. Mol. Breed. 15: 367-381.

Rohlf, F. 2000. NTSYS-PC Numerical Taxonomy and Multivariate Analysis System ver 2.11L. Applied Biostatistics, NY.

Rozen, S and H. J. Skaletsky.2000. Primer 3 on the WWW for general users and for biologist programmers, In: Krawetz S and Misener S (eds), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Pp365-386, Humana Press, Totowa, NJ.

Selvi, A., N. V. Nair, N. Balasundaram and T. Mohapatra. 2003. Evaluation of maize microsatellite markers for genetic diversity analysis and fingerprinting in sugarcane. Genome 46: 394-403.

Skinner, J.C. 1972. Selection in sugarcane: A review. Proc. Int. Soc. Sugarcane Technol. 14 : 149-162

Skinner, J. C., D. M. Hogarth and K. K. Wu . 1987. Selection methods, criteria and indices, in: Heinz D. J. (ed), Sugarcane improvement through breeding. Elsevier, Amsterdam. pp. 409-453

Sneath, P. H. A. and R. R. Sokal. 1973. Numerical Taxonomy, Freeman, San Francisco.

Sreenivasan, T. V., B. S. Ahloowalia and D. J. Heinz. 1987. Cytogenetics in: Heinz D. J. (ed), Sugarcane improvement through breeding. Elsevier, Amsterdam. pp. 211-253.

Tanksley S. D. and S. R. McCouch. (1997). Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. Sci. 277: 1063-1066.

Ude, G., M. Pillay, E. Ogundiwin and A. Tenkouano. 2003. Genetic diversiy in an African plantain core collection using AFLP and RAPD markers. Theor. Appl. Genet. 107: 248-255

Vuylsteke, M., R. Mank, B. Brugmans, P. Stam and M. Kuiper .2000. Further characterization of AFLP data as a tool in genetic diversity assessments among maize (*Zea mays* L.) inbred lines. Mol. Breed. 6:265-276

Weir, B. 1990. Genetic data analysis: methods for discrete population genetic data. Sinauer Assoc. Sunderland, Massachusetts, USA.

Whalen, M. D. 1991. Taxonomy of *Saccharum* (Poaceae). Baileya 23: 109-125

CHAPTER 3 TRAP, A NEW TOOL FOR SUGARCANE BREEDING: COMPARISON WITH AFLP AND COEFFICIENT OF PARENTAGE

3.1 Introduction

Sugarcane breeding efforts have contributed substantially to the sugar industries worldwide. In central Queensland, Australia, for example, the cultivar Q50 was nicknamed 'mortgage buster' soon after its release because of the wealth it brought to that sugar industry. In Louisiana, the popular cultivar, LCP85-384, increased cane yield by 20-25% and contributed to unprecedented boosts in sugar production. LCP85-384, like most dominant cultivars, has enjoyed widespread adoption in the Louisiana sugar industry albeit to the exclusion of other cultivars. It is well known that the over reliance on a single cultivar can result in severe consequences especially in a clonally propagated crop, such as sugarcane. Therefore, tremendous effort is being made to release new cultivars that equal or surpass the performance of LCP85-384 to Louisiana growers.

The long duration of a sugarcane selection cycle is one factor limiting the rapid development of improved sugarcane cultivars. It takes about 12 to 15 years after crossing to complete a selection cycle. Because sugarcane is clonally propagated, during this 12- to 15-year period, no new opportunities exist for sexual recombination or the creation of new genetic variation that the breeder can capitalize on. The breeder has to rely on the initial variation created during hybridization, and no amount of selection can produce a good cultivar out of a poor cross. The choice of parents to use in crossing is, therefore, one of the most crucial decisions the breeder has to make.

The complicated genome of cultivated sugarcane (high (aneu) ploidy levels, and multiple alleles at a locus) is another factor limiting progress in sugarcane breeding programs. Cultivated sugarcane was derived by mainly crossing between two species,

namely *S. officinarum* (2n = 80) and *S. spontaneum* (2n = 40 to 128), followed by backcrosses to *S. officinarum*. *Saccharum officinarum* was reported to transmit the somatic chromosome number to its F_1 progeny (Bhat and Gill, 1985; Bremer, 1961). Consequently, cultivated sugarcane harbors two genomes with about 80 % *S. officinarum* and 20 % *S. spontaneum* composition (D'Hont et al., 1996). Furthermore, chromosome numbers within cultivated sugarcane can vary (generally from about 100 to130) even among full sib progenies.

Molecular markers are valuable tools that can be used to help understand and manipulate a genome as complicated as that of sugarcane. Molecular markers can be used to tag genes for traits of economic importance such that selection for these traits (via marker-assisted selection) could occur earlier in the breeding program. Molecular markers also can be used to facilitate decisions made during crossing, as using these markers can help gain a better understanding of the genetic diversity in the parental clones. That information could then play a vital role in the utilization and management of the genotypes and indeed genes in the breeding gene pool. For example, crosses could be planned between parents from divergent backgrounds to maximize heterosis while increasing genetic diversity in the cultivated gene pool.

In sugarcane breeding programs, experimental clones are often nominated as parents based upon performance in advanced stage selection trials. Ultimately, most crosses are made among parents with phenotypic superiority in one or more key attributes with the goal of combining all key attributes in the hybrid. It is believed that the probability of recovering superior progeny is higher when both parents are themselves superior. Therefore, one would like to detect genetic diversity among phenotypically superior parents. This can be a very difficult task when relying solely on pedigree

records because superior phenotypic characteristics might have been obtained at the expense of genetic diversity at specific loci that have undergone selection. Pedigree-based estimates of genetic diversity may not account for allele frequency changes resulting from selection and genetic drift. By relying on pedigree records to estimate genetic diversity, one assumes that all genotypes are unrelated, which may be misleading in cultivated sugarcane where only a handful of clones were used in the original synthesis. Molecular markers on the other hand offer a direct comparison of genetic diversity at the DNA level without the simplifying assumptions inherent with the pedigree-based method.

When used for genetic diversity studies, molecular marker techniques, such as RFLP, RAPD, AFLP, and gSSR, customarily amplify random portions of the genome leading to competent estimates of genetic diversity. However, breeders may be more interested in results from genetic diversity studies when markers that co-segregate with traits of interest are used. However, even after quantitative trait loci (QTLs) for traits of interest have been identified, it has been argued that the underlying QTL-trait association is based on relatively large linkage blocks and could easily be lost with recombination. In addition, transferability of QTLs between populations remains a question in the minds of many plant breeders. The results from genetic diversity studies may, therefore, be more useful if the segments of the genome sampled or measured correspond to segments bearing the genes of interest to the breeder. This may be more important in sugarcane with its large genome size (estimated to be about 6 pg, approximately six times larger than that of rice) most of which may be duplicated and redundant (Ma et al., 2004).

Access to increasing numbers of sugarcane gene and expressed sequence tag (EST) sequences obtained from diverse cDNA libraries coupled with available

bioinformatics tools offer new opportunities for achieving a candidate gene approach to molecular markers in sugarcane. Target region amplification polymorphism (TRAP) is a relatively new marker technique which uses gene/EST sequence information to generate polymorphic bands around targeted/putative candidate gene regions. We previously sequenced TRAP amplicons from sugarcane and showed, using Blastx analysis, that the TRAP primers successfully amplified the anticipated candidate gene regions (Alwala et al., 2006).

Our objectives were to compare TRAP, AFLP, and pedigree-based coefficient of parentage (COP) in their ability to elucidate genetic diversity and relationships among nine sugarcane genotypes frequently used as parents in the Louisiana breeding program.

3.2 Materials and Methods

3.2.1 Plant Material and DNA Extraction

The nine sugarcane parents used in the study are described in Table 3. 1. These genotypes are experimental clones and cultivars adapted to Louisiana's unique subtropical climate. TucCP77-42, a major cultivar in northern Argentina, was bred also in Houma, LA using Louisiana adapted clones as the recurrent parents. This group of genotypes serves as an important parental pool for sugarcane crossing in Louisiana.

Young leaves were collected from each genotype, frozen immediately and stored at -80 ^{o}C. Later, the leaves were ground in liquid nitrogen. Genomic DNA was extracted using the Plant DNeasy Mini Kit (Qiagen, Valencia, CA). DNA concentrations were estimated by known concentration of lambda DNA in 1% agarose gel.

3.2.2 TRAP Markers

TRAP is a simple, 2-primer polymerase chain reaction (PCR) technique (Hu and Vick, 2003). The forward (fixed) primer was designed from genes or EST sequences and the accompanied reverse (arbitrary) primer was designed to target introns or exons. Both primers are usually about 18 bp long.

Table 3. 1. Description of the nine sugarcane genotypes used in the genetic similarity study.

Genotype	Female parent	Male parent	Comments/Agronomic characteristics
L99-238	CP79-318	LCP85-384	High sucrose parent
HoCP91-552	LCP81-10	CP72-356	High tonnage; high fiber
LCP86-454	CP77-310	CP69-380	Commercial; early high sucrose content
Ho95-988	CP86-941	US89-12	Commercial; high sugar yields and good ratooning ability
LCP85-384	CP77-310	CP77-407	Commercial; leading commercial cultivar in Louisiana from 1998 to present; high sugar yield, good ratooning ability and recently showing susceptibility to rust disease
HoCP96-540	LCP86-454	LCP85-384	Commercial; released in 2003; high sugar and cane yields; good disease resistance
HoCP95-951	CP85-866	CP85-830	BC_5 of US60-8-3; high cane yield and fiber content
TucCP77-42	CP71-321	US72-19	Commercial cultivar in Argentina; high cane yield and average sucrose content
HoCP92-624	CP81-325	CP71-1038	High sugar and cane yield; dropped due to excessive lodging; used extensively in crossing programs.

In this study, the fixed primers were designed from four genes associated with sucrose metabolism, namely sucrose synthase (SuSy), sucrose phosphate synthase (SuPS), pyruvate orthophosphate dikinase (PODK), and soluble acid invertase (SAI). The primers were designed using the Primer3 software (http://frodo.wi.mit.edu/cgi-

bin/primer3/primer3_www.cgi) (Rozen and Skaletsky, 2000). The primer optimum size, maximum size and minimum size were set to 18 nt. The optimum T_m, maximum T_m and minimum T_m were set to 53°C, 55°C, and 50°C respectively. The GenBank accession number and designed primer sequence for each gene is given in Table 2. 2 of Chapter 2.

Arbitrary reverse primer sequences were obtained from Li and Quiros (2001). The basic structure of this primer included three selective nucleotides at the 3' end, four nucleotides of AT- or GC-rich content in the core region, and 11 nucleotides as filler sequences at the 5' end. The AT and GC sequences are believed to target introns and exons, respectively. In addition, the basic rules of primer design, such as self-complementarity and maintenance of 40-60 % GC content were upheld in designing both primers (Table 2. 2 of Chapter 2). The TRAP protocol was performed on an *i-cycler* (BioRad Labs, Hercules, CA) as described in Alwala et al. (2006). After PCR, the amplified products were run on a 7 % polyacrylamide denaturing gel for 1.5 hrs at 110 W. The gel was developed and visualized using the silver staining technique. A total of 12 TRAP primer combinations were used to screen the nine parents.

3.2.3 AFLP Markers

The AFLP techinque was chosen for comparison because it has been widely used for genetic diversity studies in sugarcane (Besse et al., 1998; Lima et al., 2002) and other crops, such as beans (Bhat et al., 2005), wheat (Tian et al., 2005) and squash (Ferriol et al., 2004). Also, the power of AFLP supposedly lies in its ability to simultaneously amplify large numbers of marker loci throughout the genome (Vuylsteke et al., 2000). Thus, it was of interest to compare it to TRAP which tends to target specified regions of the genome. AFLP analysis was performed based on the protocol described by Vos et al. (1995). Two hundred nanograms of DNA were double digested with *EcoR*I and *Mse*I and

linked to specific adaptors. Primers carrying one selective nucleotide were designed, based on adaptor sequence, for pre- selective amplification. *EcoR*I and *Mse*I primers with three selective nucleotides were used for selective amplifications. All the PCR amplifications were carried out on an *i-cycler* (BioRad Labs, Hercules, CA). The amplified products were mixed with equal amount of dye (composition) and 5 µl of each sample was separated by electrophoresis on a 6 % polyacrylamide denaturing gel for 2 hr at 110 W. The gels were documented using the silver staining technique. A total of 28 *EcoR*I /*Mse*I AFLP primer combinations were used to screen the nine parents.

3.2.4 Estimation of TRAP- and AFLP-derived Genetic Diversity and Polymorphic Information Content

The bands from TRAP- and AFLP-derived gels were scored as '1' for presence and '0' for absence. Jaccard-similarity coefficient (1908) was used to estimate genetic diversity (GS) between pairs of genotypes as follows: $GSij = a/(a+b+c)$, where *GSij* is the genetic similarity measurement between individuals *i* and *j*, *a* represents the number of matching bands present in both individuals whereas *b* and *c* are the number of bands present in individual *i* and *j,* respectively, but not in their counterparts. The bands absent in pairs of individuals were excluded from the calculation.

Allelic diversity at a given locus can be measured by polymorphism information content (PIC) wherein a marker can distinguish two alleles taken at random from a population and it was calculated as follows:

$$PIC = 1- \Sigma f_i^2$$

where, f_i is the frequency of the i^{th} allele (Weir 1990). Considering the number of alleles at a locus along with their relative frequencies in a given population, an estimate of the

discriminatory power of a marker can be obtained by calculating the PIC value (Vuylsteke et al., 2000).

3.2.5 Coefficient of Parentage

The coefficient of parentage (COP), which corresponds to the probability that alleles at a locus in two individuals are identical by descent, was calculated to represent the pedigree-based measure of genetic diversity. The COP was calculated based on Kempthorne (1957) using the PROC INBREEDING procedure in SAS (SAS Inc., 2002). The COP value between remotely related parents was assumed to be 0, and each genotype was assumed to receive half of their genome from each of its parent. All of the ancestors were assumed to be heterozygous, since sugarcane is a highly heterozygous crop, and in addition, the COP of a genotype with itself was assumed to be 0.5 rather 1.0 as for homozygous inbreds like rice (Kempthorne, 1957; Chiang and Lo, 1993; Deren, 1995).

3.2.6 Cluster and Principal Coordinate Analyses

For ease of interpretation, the genetic similarity (GS) values for TRAP, AFLP, and COP between pairs of genotypes were subjected to both cluster (CA) and principal coordinate (PCoA) analyses to obtain graphical representations of the relationships between the nine genotypes. The goodness of fit of the dendrograms formed from the GS matrix was evaluated by means of the cophenetic coefficient of correlation. A minimum-length spanning tree (MST) was superimposed on the PCoA plot to help detect local distortion because pairs of points which look close together in a plot may actually be far apart if other dimensions are taken into account. These analyses were performed using the NTSYS-PC ver 2.2 (CA; Rohlf, 2000) and PAST (PCoA; Hammer et al., 2001) software packages. Bootstrap analysis with 1000 replications with a 50 % consensus rule

was performed using the PAUP ver 4.0 software (Sinauer Associates, Inc., MA), and the boostrap values were superimposed on the CA dendrogram as a measure of the robustness of branches on the dendrogram.

3.2.7 Correlation between COP, TRAP and AFLP-derived GS

The correlation among pairs of the three genetic diversity measures was compared using two methods. The first method employed the MAXCOMP routine of NTSYS-PC software, in which two GS matrices are compared by estimating the normalized Mantel Z-statistic (Mantel, 1967). The second method estimated the simple or Pearson's correlation coefficient (*r*) between the measures.

3.2.8 Bootstrap Analysis

Bootstrap analysis (Efron, 1982) was carried out to investigate if the number of markers used to generate GS were sufficient to provide precise estimates among the genotypes. Subsamples, consisting of different number of polymorphic bands, were generated by re-sampling 1000 times, with replacement, to estimate GS between every two pairs of genotypes for each subsample. The average coefficient of variation was estimated across subsamples for a given number of polymorphic bands. The analysis was performed using the Dboot software, kindly provided by Dr. A. S. G. Coelho (Universidade Federal de Goias/Goiana-GO).

3.3. Results

3.3.1 TRAP Markers

3.3.1.1 Percent Polymorphism and PIC Values

All 12 TRAP primer combinations produced multiple PCR fragments (bands) in each of the nine cultivars which ranged in size from 300 to 700 bp (Fig. 3. 1). A total of 444 unambiguous bands were scored, of which 242 (55 %) were polymorphic (Table 3.

2). The total number of bands amplified by individual primer combinations ranged from 19 (SuPS + Arbi 3) to 69 (SuSy + Arbi 1) with an average of 37 bands per primer combination. These two primer combinations were also responsible for the least (10) and most (41) number of polymorphic bands produced with an average of 20 polymorphic bands per primer combination. The PIC values, averaged over all polymorphic loci for individual primer combinations, varied from 0.32 to 0.40 with an overall mean of 0.36 ± 0.12. From the polymorphism produced, it was possible to distinguish each of the nine genotypes. Sometimes only one of the primer combinations was needed to distinguish all nine genotypes. Bands were found to be uniquely present or absent in some genotypes. TRAP fragments were found to be highly reproducible (Fig. 3. 1).

3.3.1.2 Genetic Diversity and Relationship among Genotypes

The TRAP-GS mean estimate, averaged across all pair-wise combinations of genotypes, was generally high (0.75 ± 0.04) as expected based on the shared ancestry among these genotypes (Table 3. 1, Fig. 3. 2). The GS values ranged from 0.67 (between HoCP91-552 and HoCP92-624) to 0.87 (between LCP85-384 and HoCP96-540) (Table 3. 4, Fig. 3. 3A). Data from the GS matrix were visualized using two methods, CA and PCoA.

The CA yielded a dendrogram (Fig. 3. 3A) with a cophenetic coefficient of correlation value of 0.81, which is above the 0.80 generally regarded as a good fit (Rohlf and Sokal, 1981). No distinct clusters were found. However, a subgroup was apparent between LCP85-384, HoCP96-540, and LCP86-454 (Fig. 3, 3A), which is in agreement with the close relationship known among these genotypes (Table 3. 3). Missing from this subgroup, however, was L99-238, a progeny of LCP85-384. The average GS between members of the group containing LCP85-384, HoCP96-540 and LCP86-454 vs. Ho95-

988 is 0.78 while for L99-238 it is 0.77. This might explain why using the clustering technique placed Ho95-988 closer to the group compared with L99-238.

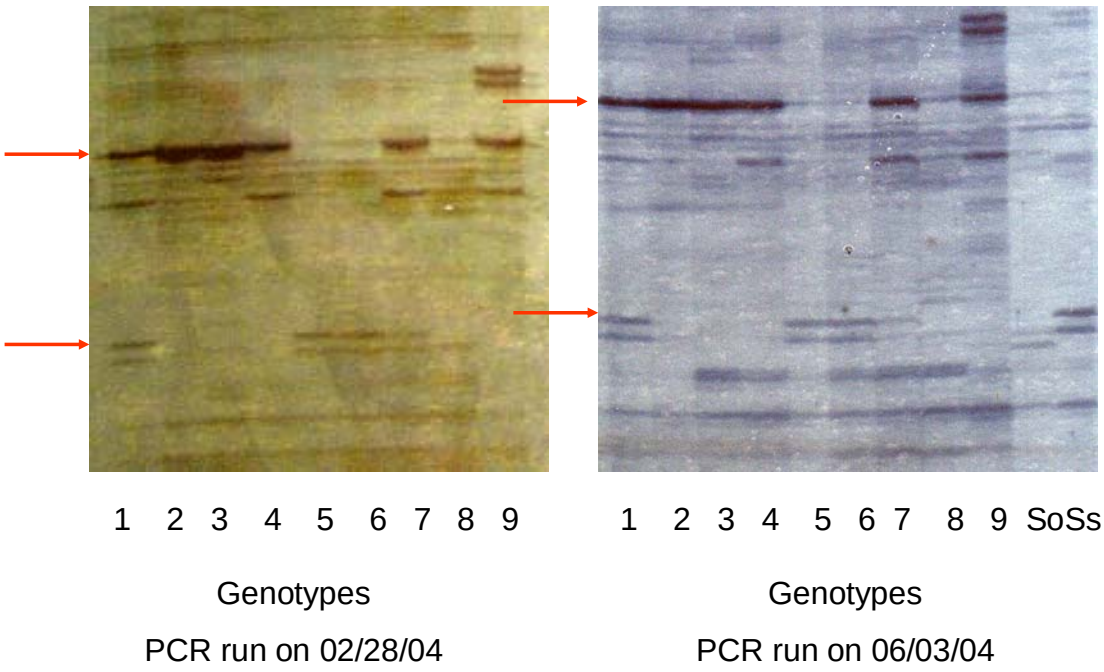


Figure 3. 1. Reproducibility of TRAP markers depicted here by segments from two silver-stained polyacrylamide gels with polymorphic TRAP fragments generated using SuSy + Arbi 3 (see Table 3. 1). Genotypes 1 to 9 are similar on the two gels with the reactions and gels run on different dates. Genotypes: 1= L99-238; 2 = HoCP91-552; 3 = LCP86-454; 4 = Ho95-988; 5 = LCP85-384; 6 = HoCP96-540; 7 = HoCP95-951; 8 = TucCP77-42; 9 = HoCP92-624; So = La Stripe; Ss = SES 147 B. Arrows show identical banding patterns between the two gels.

The bi-plot from PCoA superimposed with the MST portrayed a slightly different sub-grouping which was composed of LCP85-384 and its two progeny, L99-238 and HoCP96-540 (Fig. 3. 3B, Table 3. 4). Considering the first principal coordinate, LCP86-454, a sibling of LCP85-384, was placed outside of the subgroup (Fig. 3. 3B) although

the MST clearly illustrates the relationship between the two genotypes. Both the CA and PCoA portrayed HoCP91-552 as the genotype most distant from the rest of the group (Figs. 3. 3A and 3. 3B). In the dendrogram (Fig. 3. 3A), the split of HoCP91-552 from the rest of the genotypes was only one of two branches supported by a bootstrap value greater than 50%.

Table 3. 2. Percent polymorphism and PIC values of TRAP markers used in genotyping nine sugarcane parents from the Louisiana breeding program.

	Primer combinations	No. bands observed	No. polymorphic bands	Percent polymorphism	PIC†
1	SuSy + Arbi 1	69	41	59.42	0.33
2	SuSy + Arbi 2	60	19	31.66	0.38
3	SuSy + Arbi 3	54	38	70.37	0.35
4	SuPS + Arbi 1	25	15	60.00	0.39
5	SuPS + Arbi 2	45	24	53.33	0.38
6	SuPS + Arbi 3	19	10	52.63	0.40
7	PODK + Arbi 1	31	14	45.16	0.32
8	PODK + Arbi 2	28	18	64.28	0.40
9	PODK + Arbi 3	34	15	44.11	0.35
10	SAI + Arbi 1	35	21	60.00	0.36
11	SAI + Arbi 2	20	13	65.00	0.32
12	SAI + Arbi 3	24	14	58.33	0.34
Total		444	242		
Average		37	20	55	0.36

† Polymorphism information content

3.3.1.3 Bootstrap Analysis

As expected, the precision (CV %) with which one can distinguish among the nine genotypes increased with increasing numbers of polymorphic TRAP bands (Fig. 3. 4). Using all 242 polymorphic TRAP bands, the CV for distinguishing among the nine

genotypes in the study was 8.1 % (Fig. 3. 4), which seems to be reliable considering some authors recommend a CV of 10 % (Lima et al., 2002). About 160 polymorphic TRAP bands would be necessary to measure GS among the nine genotypes with 10 % precision.

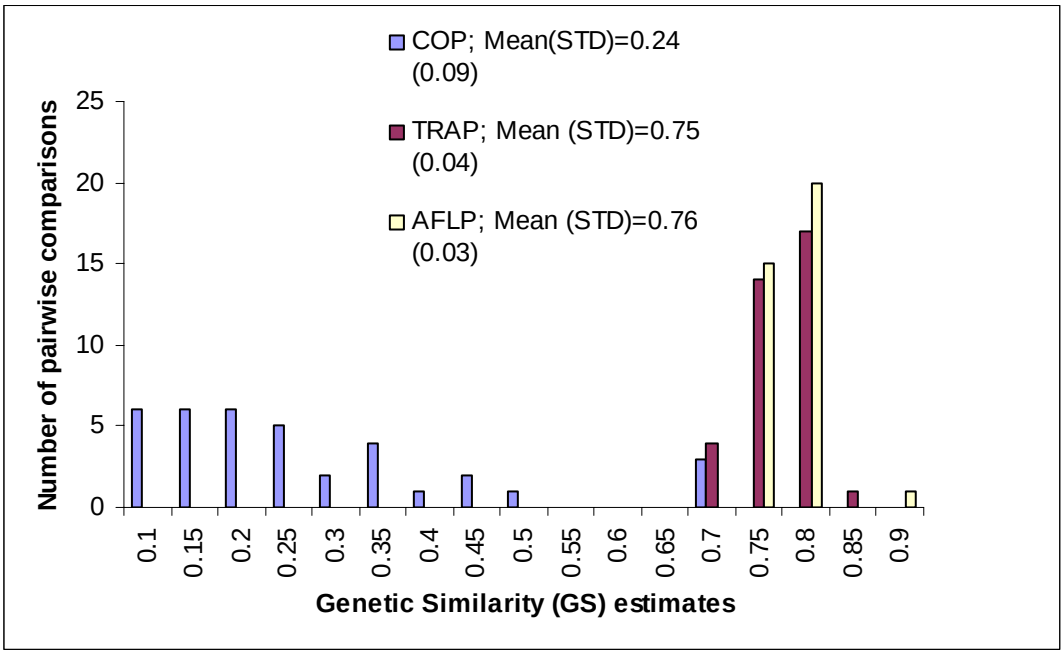
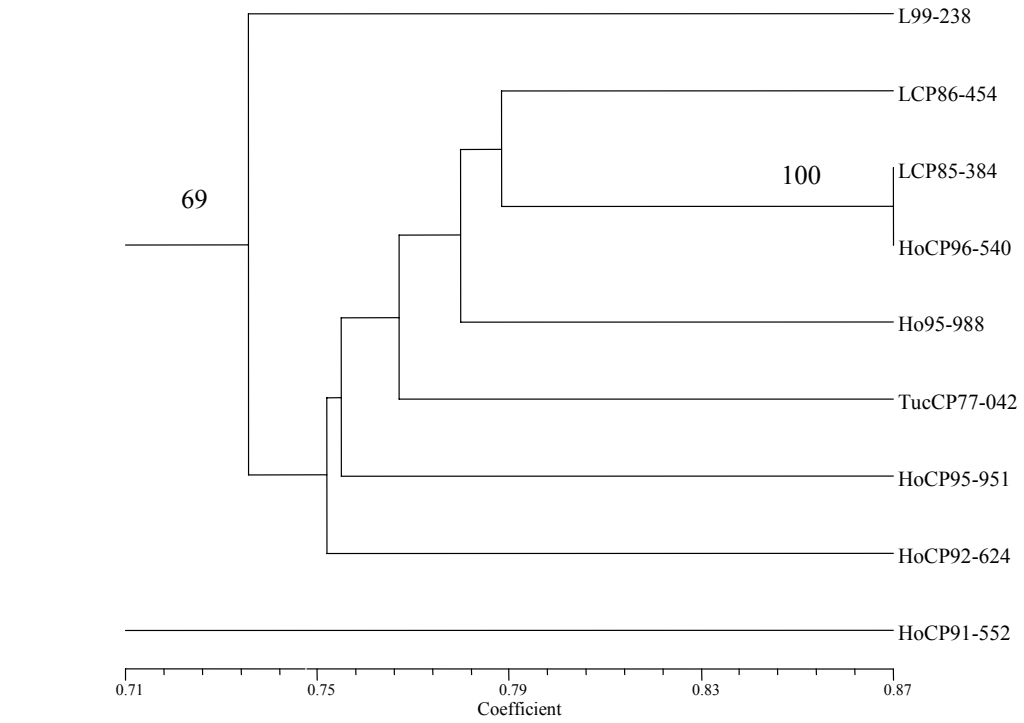


Figure 3. 2. Frequency distribution of genetic similarity estimates based on pedigree (COP), TRAP, and AFLP data. Note: The theoretical range of genetic similarity (GS) values for molecular markers is 0 to 1, whereas COP values in a heterozygous crops, e.g. sugarcane, range from 0 to 0.5. To facilitate comparison among the different GS estimates, the COP values were multiplied by 2.

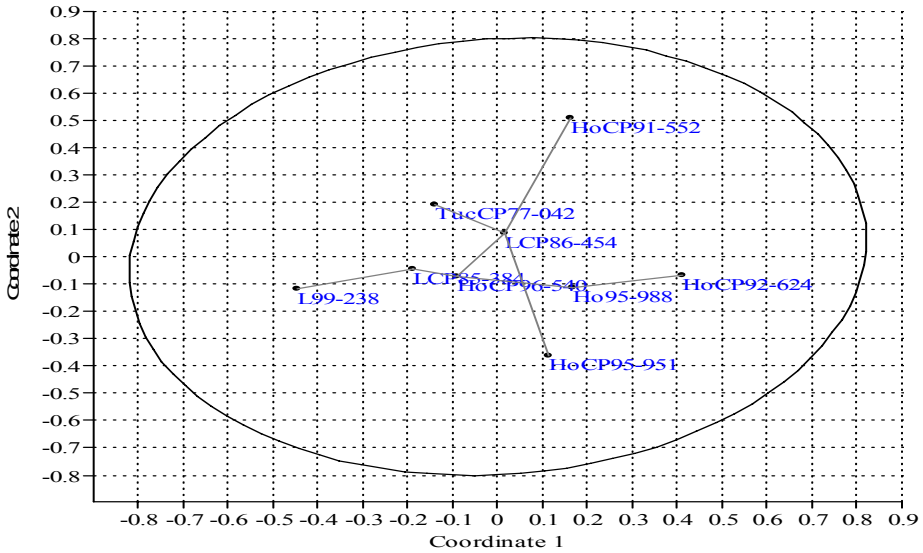
3.3.2 AFLP Markers

3.3.2.1 Percent Polymorphism and PIC Values

A total of 40 AFLP primer combinations were tested of which 28 primer combinations were adequate to study genetic relationships within this set of sugarcane



A



B

Figure 3. 3. Jaccard's genetic similarity pattern among nine sugarcane genotypes based on TRAP polymorphism and depicted by cluster analysis (A) and principal coordinate analysis (B). The numbers on the dendrogram (A) represent 50 % majority rule bootstrapped values.

genotypes. The 28 primer combinations produced a total of 1325 bands of which 686 (53%) were polymorphic (Table 3. 4). The unambiguous bands ranged in size from 250 to 600bp. The total number of bands per primer combination ranged from 23 (E-AAC+M- CTC) to 81 (E-ACA+M-CTC) with an average of 47 bands per primer combination. These two primer combinations were also responsible for the least (8; E-

Table 3. 3. Pairwise genetic similarity (GS) estimates using TRAP and AFLP markers compared with coeeficient of parentage (COP) on nine parents.

TRAP	L99-238	HoCP91-552	LCP86-454	Ho95-988	LCP85-384	HoCP96-540	HoCP95-951	TucCP77-042	HoCP92-624
L99-238	1.00								
HoCP91-552	0.68	1.00							
LCP86-454	0.76	0.78	1.00						
Ho95-988	0.72	0.71	0.78	1.00					
LCP85-384	0.79	0.73	0.79	0.78	1.00				
HoCP96-540	0.78	0.71	0.79	0.79	0.87	1.00			
HoCP95-951	0.69	0.71	0.77	0.75	0.75	0.76	1.00		
TucCP77-042	0.72	0.72	0.79	0.74	0.78	0.77	0.76	1.00	
HoCP92-624	0.70	0.67	0.76	0.77	0.75	0.75	0.76	0.75	1.00
AFLP									
L99-238	1.00								
HoCP91-552	0.78	1.00							
LCP86-454	0.79	0.8	1.00						
Ho95-988	0.73	0.78	0.76	1.00					
LCP85-384	0.80	0.77	0.75	0.76	1.00				
HoCP96-540	0.77	0.75	0.79	0.73	0.77	1.00			
HoCP95-951	0.76	0.75	0.75	0.76	0.74	0.77	1.00		
TucCP77-042	0.75	0.74	0.76	0.72	0.75	0.74	0.75	1.00	
HoCP92-624	0.77	0.74	0.73	0.72	0.84	0.79	0.72	0.74	1.00
COP									
L99-238	0.50								
HoCP91-552	0.10	0.50							
LCP86-454	0.15	0.08	0.50						
Ho95-988	0.05	0.04	0.03	0.50					
LCP85-384	0.35	0.08	0.18	0.04	0.50				
HoCP96-540	0.25	0.08	0.35	0.04	0.35	0.50			
HoCP95-951	0.20	0.11	0.11	0.06	0.20	0.15	0.50		
TucCP77-042	0.07	0.08	0.06	0.03	0.07	0.07	0.08	0.50	
HoCP92-624	0.15	0.10	0.11	0.04	0.14	0.13	0.16	0.09	0.50

AAC+M-CTC) and the most (44; E-ACA+M-CTC) number of polymorphic bands produced, with an average of 24 polymorphic bands per primer combination. The PIC

value, averaged over all polymorphic loci for individual primer combinations, varied from 0.27 (E-ACA+M-CAC) to 0.45 (E-ACA+M-CAG) with an overall mean 0.35 ± 0.12. As with TRAP markers, it was possible to distinguish each one of the nine genotypes. Sometimes just one AFLP primer combination was sufficient to distinguish all the genotypes. Also, bands were found to be uniquely present or absent in some genotypes.

3.3.2.2 Genetic Diversity and Relationship among Genotypes

The AFLP-GS estimates between pairs of genotypes ranged from 0.72 (between HoCP92-624 and Ho95-988) to 0.84 (between HoCP92-624 and LCP85-384) with a mean value of 0.76 ± 0.03 (Table 3. 3, Fig. 3. 2). Cluster analysis produced a dendrogram with a cophenetic coefficient of correlation value of 0.75 (Fig. 3. 5A). The dendrogram had two distinct clusters although the bifurcation had only marginal (45 %) bootstrap support. LCP85-384 was placed in a cluster with both of its progenies, L99-238 and HoCP96-540, but not with its half sibling LCP86-454 (Fig. 3. 5A; Table 3. 1). Surprisingly, the closest and most robust (100 % bootstrap support) relationship was found between LCP85-384 and HoCP92-624, a genotype with which it seemingly does not share a recent lineage (Table 3. 1). Similar results were depicted by the PCoA with the same two groups of genotypes apparent in the first principal coordinate (Fig. 3. 5B).

3.3.2.3 Bootstrap Analysis

A total of 1325 polymorphic bands were revealed by AFLP markers. However, for comparison with the 242 polymorphic bands revealed by TRAP markers, each of 10 subsamples (with replacement) of 250 polymorphic AFLP bands was subjected to bootstrap analysis. Based on the 250 polymorphic AFLP bands, GS was measured among the nine genotypes with 8.3 % precision (Fig. 3. 4). About 160 polymorphic

AFLP bands would be necessary to measure GS among the nine genotypes with 10 % precision.

3.3.3 COP

3.3.3.1 Genetic Diversity and Relationship among Genotypes

Since sugarcane is highly heterozygous, the COP-GS estimate of a genotype with itself was assumed to be 0.5 and 0 between remotely related parents. Therefore, unlike in rice for example, the COP estimates in sugarcane range from 0 to 0.5 rather than from 0 to 1. The COP-GS estimates among the nine genotypes varied from 0.03 (between TucCP77-42 and Ho95-988 and Ho95-988 and LCP86-454) to 0.36 (between HoCP96-

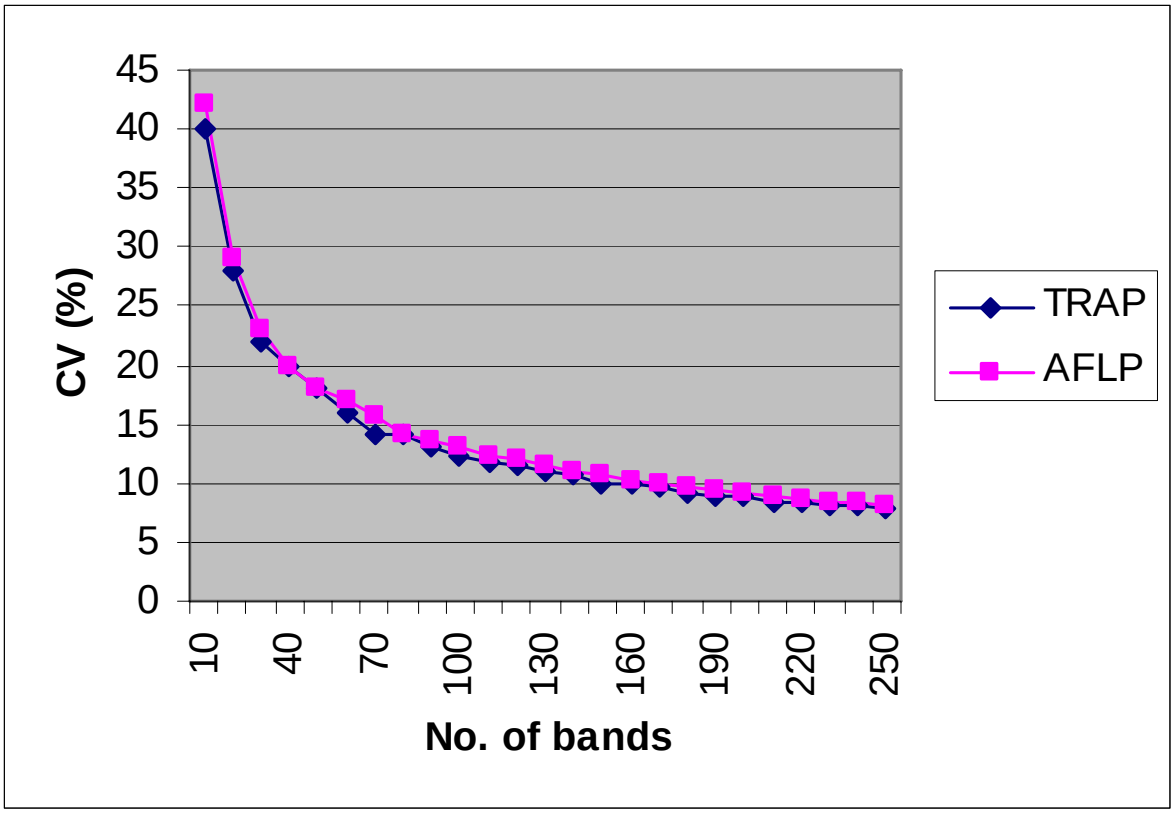


Figure 3. 4. Number of polymorphic TRAP and AFLP bands necessary to estimate genetic similarity among nine sugarcane genotypes with a given level of precision (CV %) as estimated using bootstrap analysis.

540 and LCP85-384) with a mean of 0.12 ± 0.09 (Table 3. 3; Fig. 3. 2). Cluster analysis of the pair-wise COP matrix resulted in a dendrogram with a cophenetic coefficient of correlation value of 0.92 (Fig. 3. 6A). The COP-derived dendrogram revealed no distinct pattern of diversity. For example, although one could still trace the relationship among

Table 3. 4. AFLP primer combinations and the number of total and polymorphic bands observed in a set of nine sugarcane genotypes.

	Primer combination	Bands observed	Polymorphic bands	Percent polymorphism	PIC†
1	E-ACT+M-CAT	64	28	43.75	0.33
2	E-ACT+M-CAA	37	15	40.54	0.33
3	E-ACT+M-CTC	36	35	97.22	0.38
4	E-ACT+M-CTG	53	29	54.71	0.38
5	E-AAC+M-CAA	64	21	32.81	0.35
6	E-AAC+M-CTA	44	26	59.09	0.37
7	E-AAC+M-CTC	23	8	34.78	0.36
8	E-AAC+M-CTG	66	30	45.45	0.39
9	E-ACC+M-CAA	53	17	32.07	0.37
10	E-ACC+M-CTA	55	23	41.81	0.32
11	E-ACC+M-CTC	50	19	38.00	0.33
12	E-ACC+M-CTG	53	22	41.51	0.33
13	E-ACA+M-CAA	39	22	56.41	0.37
14	E-ACA+M-CTA	40	31	77.50	0.29
15	E-ACA+M-CTC	81	44	54.32	0.35
16	E-ACA+M-CTG	68	23	33.82	0.35
17	E-AGC+M-CAT	25	20	80.00	0.35
18	E-AGC+M-CAA	53	18	33.96	0.35
19	E-AGC+M-CTG	46	22	47.82	0.32
20	E-ACG+M-CAT	40	24	60.00	0.39
21	E-ACG+M-CAA	44	27	61.36	0.35
22	E-AAC+M-CAC	45	25	55.55	0.35
23	E-ACC+M-CAC	49	26	53.06	0.39
24	E-ACA+M-CAC	37	26	70.27	0.27
25	E-AGC+M-CAC	40	23	57.50	0.37
26	E-ACC+M-CAG	44	29	65.90	0.36
27	E-ACA+M-CAG	35	21	60.00	0.45
28	E-AGC+M-CAG	41	32	78.04	0.37
	Total	1325	686		
	Average	47	24	53	0.35

† Polymorphism information content

genotypes, such as LCP85-384, HoCP96-540, L99-238 and LCP86-454, which are known to share a common lineage (Table 3. 1), CA displayed no dichotomy between this group (related) and the non-related genotypes in the study. This dichotomy was clearly revealed by the PCoA-derived bi-plot (Fig. 3. 6B).

3.3.4 Associations between Pedigree, TRAP, and AFLP-derived GS Estimates

Similar levels of association, as estimated with both the normalized Mantel Z-statistic and Pearson correlation coefficient (Table 3. 5), were found between COP-GS with TRAP-GS and between COP-GS with AFLP-GS. Although the correlation values were significant ($P < 0.05$; N = 36) they were moderate (0.40). Both methods (Mantel and Pearson) also calculated a similar level of correlation between the two (TRAP and AFLP) molecular marker-derived GS estimates. The correlation between TRAP and AFLP-GS estimates was much lower (0.14) compared with their respective correlations with the COP-GS (0.40).

3.4 Discussion

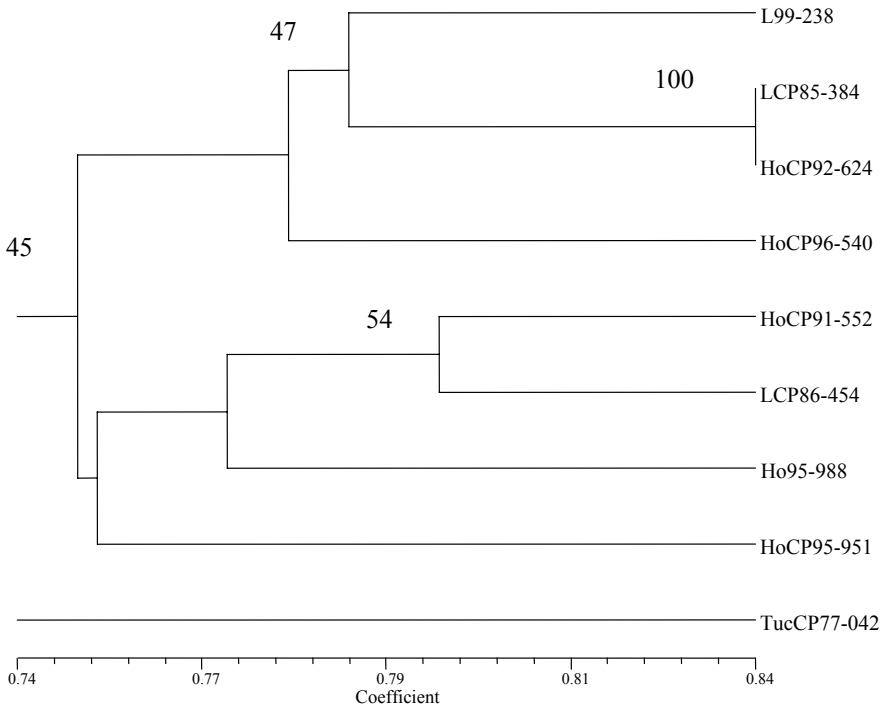
3.4.1 Comparing Characteristics of TRAP and AFLP Markers

The values of percent polymorphism and PIC reported in this study are typical for sugarcane. Using 21 AFLP primer combinations, Lima et al. (2002) detected an average of 50 % polymorphism among 79 Brazilian sugarcane cultivars, whereas Selvi et al. (2005) reported an average of 52 % among 28 cultivars from India using 12 primer combinations. Selvi et al. (2005) reported PIC values for AFLP that ranged from 0.31 to 0.41. TRAP analysis based upon 24 primer combinations among 61 sugarcane genotypes from Canal Point, Florida detected 58 % polymorphism and a PIC value of 0.32 (Edme, personal communication). The complex polymorphism profile displayed in sugarcane for AFLP (Besse et al., 1998; Lima et al., 2002; Selvi et al., 2005; this study) and TRAP

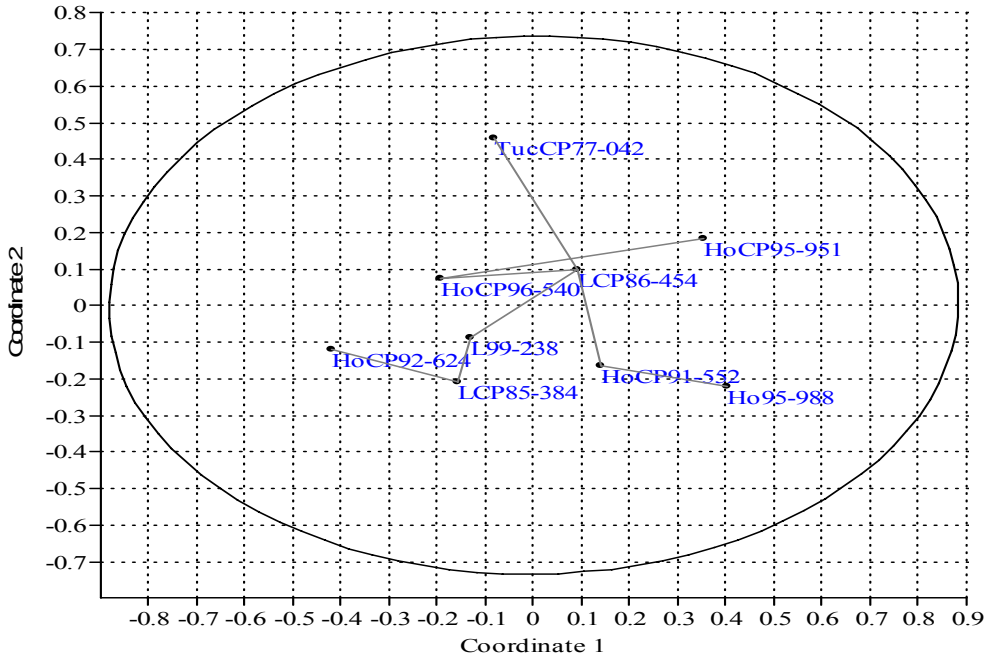
(Arro, 2004; Edme, personal communication; this study) markers can be attributed to its large genome size, high levels of heterozygosity and aneuploidy, which is perpetuated via vegetative propagation. No two of the nine genotypes presented an identical profile, indicating that TRAP markers, besides AFLP, can be useful for sugarcane fingerprinting.

The utility of a DNA marker technique can be defined by its multiplex ratio (number of markers that can be generated in one single reaction) and the PIC (effective number of alleles that can be detected per marker in a set of individuals, i.e., the discriminatory power of the marker) (Powell et al., 1996; Vuylsteke et al., 2000). The ability of the AFLP technique to simultaneously amplify a large number of marker loci throughout the genome has been cited as a major advantage of AFLPs over other marker systems (Vuylsteke et al., 2000). In this study, AFLP was only marginally superior to TRAP with regards to the total number of bands amplified per primer combination. However, similar PIC values were found between the two marker systems. In addition, similar numbers of polymorphic bands were necessary to distinguish among the nine genotypes with 10 % precision. The overall percent polymorphism was somewhat higher for TRAP (55 %) than for AFLP (53 %). Thus, on the basis of these data, a similar level of polymorphism detection efficiency is to be expected from these two dominant markers.

Experience in our lab has shown, however, that the relative polymorphism detection efficiency between AFLP and TRAP may be dependent upon the genetic structure of the population under study. For example, different results were obtained when the same set of TRAP and AFLP markers were used to genotype 100 individuals from an interspecific (*Saccharum officinarum* 'La Stripe' x *S. spontaneum* 'SES 147B') mapping population. The total number of bands amplified and percent polymorphism

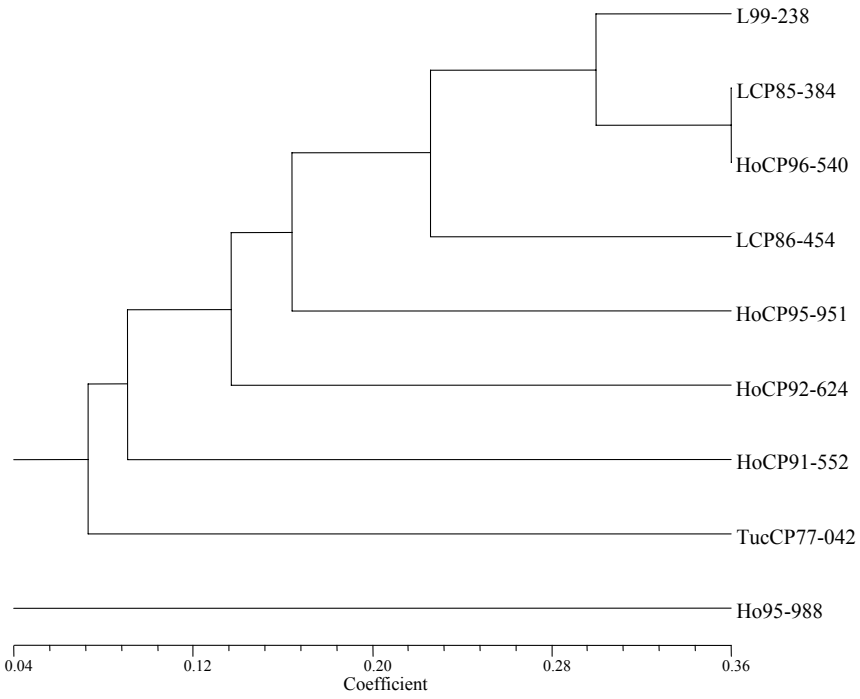


A

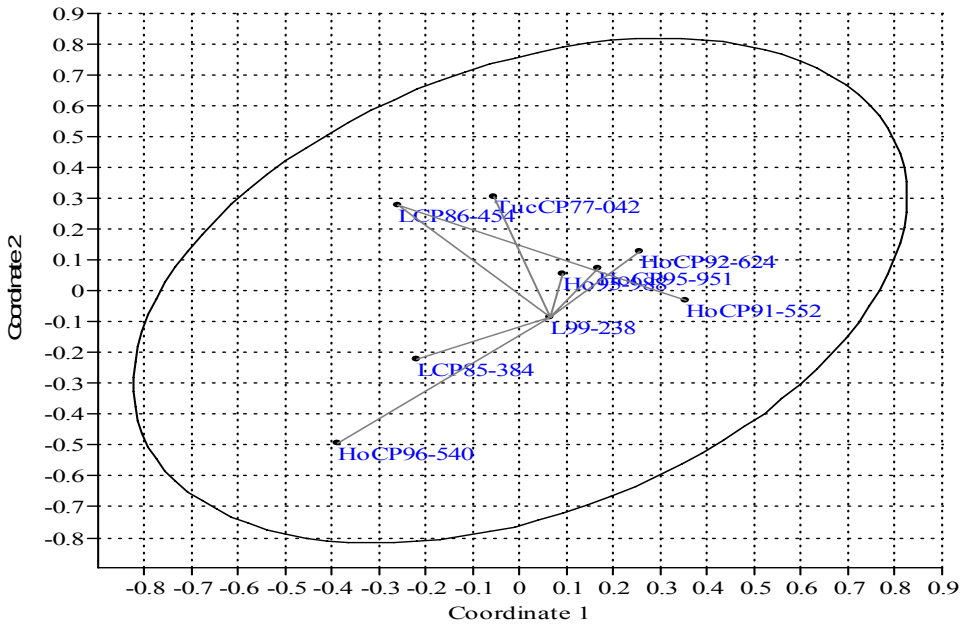


B

Figure 3. 5. Jaccard's genetic similarity pattern among nine sugarcane genotypes based on AFLP polymorphism and depicted by cluster analysis (A) and principal coordinate analysis (B). The numbers on the dendrogram (A) represent 50 % majority rule bootstrapped values.



A



B

Figure 3. 6. Genetic similarity pattern among nine sugarcane genotypes based on the coefficient of parentage and depicted by cluster analysis (A) and principal coordinate analysis (B).

Table 3. 5. The Mantel's Z-statistic and Pearson correlation coefficient for pair-wise comparisons between the COP, TRAP and AFLP-derived estimates of genetic similarity (GS).

	Normalized Mantel Z statistic *R*	Pearson correlation coefficient (*r*)
AFLP vs COP	0.41	0.42
TRAP vs COP	0.40	0.41
AFLP vs TRAP	0.14	0.14

revealed by AFLP surpassed that of TRAP by about three to four fold (unpublished data) in the La Stripe x SES 147B population. Because TRAP primers are designed to target only a small and specified portion of the genome (Hu and Vick, 2003; Alwala et al.,2006), AFLP markers may be more robust for detecting polymorphism among closely related genotypes, as they are more likely to sample different segments throughout the genome. In soybean, Powell et al. (1996) found good correlations between AFLP and other markers (RFLP, RAPD and SSR) at the interspecies level,which disappeared at the intraspecies level, with AFLP giving the best resolution among genotypes. It is best to allow research objectives to guide the decision of choosing the appropriate DNA marker technique(s).

3.4.2 Comparing GS Estimates

The mean, range, and distribution values for TRAP-GS and AFLP-GS were similar, but both were distinct from COP-GS (Table 3. 4; Fig. 3. 2). The mean values from TRAP-GS and AFLP-GS highlight the narrow genetic base reported for cultivated sugarcane (Mangelsdorf, 1983; Deren, 1995; Lima et al., 2002; Arro, 2005). However, judging from the COP-GS, it would appear that substantial amounts of genetic diversity exist in sugarcane. Up to 55 % (20/36) of the COP-GS were below 0.1, suggesting that only about 45 % of the genetic material segregating in the ancestral population was identical by descent between any two genotypes in this study (Table 3. 4). In calculating

TRAP-GS and AFLP-GS, only polymorphic bands were taken into consideration, yet lower levels of genetic diversity (high GS) were detected by these methods compared with the COP method. Moreover, of the total bands amplified by TRAP and AFLP markers, 45 % and 47 %, respectively, were monomorphic and therefore identical in state. This tendency of the COP method to overestimate genetic diversity compared to DNA-based methods has been reported by other authors (Cox et al., 1985; Barbosa-Neto et al., 1996; Kim and Ward, 1997).

As with several previous studies (Cox et al., 1985; Graner, 1994; Barbosa-Neto et al., 1996; Kim and Ward, 1997; Sun et al., 2003) this study found moderate levels of association between the DNA- and COP-based estimates of GS. In wheat, a low *r* value of 0.27 was observed by Cox et al. (1985) between isozyme-based GS and COP. RFLP-based GS with COP in barley generated a low correlation value of 0.27 for winter type and a moderate value of 0.42 for spring type (Graner et al., 1994). Evaluating the correlation between RAPD-based GS with COP resulted in a low *r* value of 0.10 in potatoes (Sun et al., 2003). This disparity stems from the fact that the assumptions inherent in calculating COP are unrealistic for most cultivated species, and sugarcane is no exception (Deren, 1995). For example, the COP method assumes that both parents contribute equally (half of their alleles) to the offspring, essentially ignoring the effect of selection and genetic drift during cultivar development. As evident from Table 3. 4, the relationship between LCP85-384 and its two progenies HoCP96540 and L99238 was not equal for TRAP-GS and AFLP-GS. Furthermore, it is well known that chromosome numbers within cultivated sugarcane can vary (generally from about 100-130) even among full-sib progenies. This can substantially affect DNA-based measurements of GS, but is yet unaccounted for by currently available models for estimating COP.

Considering that only a handful of clones were used in the original nobilization event to derive modern sugarcane, the assumption that two clones in this study are unrelated (COP = 0) relative to the original ancestors would be unrealistic. Thus, as opposed to the DNA-based methods, the COP method cannot account for alleles that are alike in state but not identical by descent resulting in a disproportionate downward bias of GS estimates. Incomplete pedigree records or errors in annotating parents would help to over emphasize the downward bias of COP estimates. Moreover, selfing could distort pedigree records and crossing with stray pollen could bias COP. To minimize this bias, we recalculated the correlation coefficients after eliminating COP values < 0.1. The correlation between COP-GS and TRAP-GS increased (r = 0.69; N = 16), that for COP-GS with AFLP-GS decreased (r = 0.31; N = 16), while that between TRAP-GS and AFLP-GS remained unchanged (r = 0.16; N =16). However, when values for the three closest known relatives were removed (i.e. COP = 0.35) the correlations decreased to 0.06 (COP-TRAP), 0.22 (COP-AFLP, and 0.11 (TRAP-AFLP). The lack of congruence and consistency among TRAP-GS, AFLP-GS, and COP-GS, throughout the range of diversity detected among the genotypes in this study, suggests that the three measures detect different aspects of relatedness.

Several authors (Graner, 1994; Barbosa-Neto et al., 1996; Kim and Ward, 1997; Sun et al., 2003) have recommended molecular marker-based estimates of GS to be more reliable than COP. This is largely because molecular markers such as TRAP and AFLP directly measure DNA sequence variations. However, a drawback of markers, such as TRAP and AFLP, is that the utility of bands produced by these markers can be confounded by lack of locus specificity. Without sequencing, it would be difficult to state unequivocally that bands or alleles that are identical in state (i.e. migrating to the

same position on a gel) are not co-migrating non-homologous bands. Lack of adequate genome coverage is another factor that can limit the utility of DNA-based estimates of GS. This can be resolved by using markers for which the genome location is known such that markers that span the entire genome are chosen.

3.4.3 Comparing Genetic Diversity Patterns among Genotypes

Following CA, the least distinct pattern was obtained from the COP dendrogram while AFLP gave the most distinct pattern. However, it was easier to explain the TRAP dendrogram based on pedigree records. The genetic resolution and interpretation of the data was enhanced by including the PCoA bi-plots. In general, when the dendrogram and bi-plot were considered together, the three measures seemed to depict a somewhat similar pattern of relationship among the genotypes, the major exception being the tight relationship (100 % bootstrap support) between LCP85-384 and HoCP92-624. A closer examination of the pedigree tree revealed that the maternal grandparents of LCP85-384 (CP52-068 x L65-69) were indeed the great grandparents of HoCP92-624 (CP52-068 x CP62-258) x (CP65-357 x L65-69). One could speculate that AFLP markers may be detecting favorable alleles or blocks of genes from these ancestral parents that were preserved through independent selection for the same trait(s) in the two cultivars. As for the TRAP markers, only sucrose related primers were exemplified in this study which may not be identical to the ones detected here by AFLP markers.

3.5 Conclusions

The results showed that TRAP markers have utility for sugarcane genetic diversity studies. TRAP markers produced percent polymorphism and PIC values similar to that of AFLP markers and measured GS with the same level of precision as AFLP markers. A similarly moderate level of association was found between TRAP-GS and

COP-GS estimates and between AFLP-GS and COP-GS estimates. The association between TRAP-GS and AFLP-GS was low. Violations of the assumptions used in calculating COP was partly responsible for the moderate level of association between COP and the two DNA-based estimates, as the COP method tends to underestimate GS. However, exclusion of subsets of data along the range of COP-GS estimates led to different levels of association between COP and TRAP, COP and AFLP and TRAP and AFLP suggesting that the three measures could be detecting different aspects of GS. Notwithstanding, with few exceptions, the dendrograms and bi-plots produced using the three measures depicted a somewhat similar pattern of diversity among the genotypes. Therefore, some combination of TRAP, AFLP and COP would likely be more useful in estimating GS, as this would compensate for the inaccuracies inherent within each of the methods.

Estimates of GS could be incorporated as a tool to assist sugarcane breeders with selecting the most divergent parents to maximize heterosis and transgressive segregation in the progeny population. The inexpensive COP could be used as a first step to assemble a large diverse group of potential parents. Molecular markers, such as TRAP and AFLP, could then be used to confirm the pedigrees. Moreover, molecular markers provide a more direct and precise estimate of allele frequency differences among the parents. Decisions could be made to decide the best crosses based on the GS values among the parents, thus allowing the breeder to focus attention and resources on the most promising crosses. Only loci for which the parents carry different alleles are expected to contribute to genetic variance in the progeny population. If such loci co-locate with genes governing the traits being measured, then it may be possible to predict hybrid performance based on GS among parents.

3.6 References

Alwala, S., A. Suman, J.A. Arro, J.C. Veremis, and C.A. Kimbeng. 2006. Target Region Amplification Polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. Crop Sci. 46:448–455.

Arro, J. A. 2004. Genetic diversity among sugarcane clones using Target Region Amplification Polymorphism (TRAP) markers and pedigree relationships. Masters Thesis submitted to Louisiana State University, Baton Rouge, LA.

Barbosa-Neto, J.F., M.E. Sorrells, and G. Cisar. 1996. Prediction of heterosis in wheat using coefficient of parentage and RFLP-based estimates of genetic relationship. Genome. 39:1142-1149.

Besse, P., G. Taylor, B. Carroll, N. Berding, D. Burner, and C.L. McIntyre. 1998. Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. Genetica 104:143–153.

Bhat, K.V., S. Lakhanpal, and S. Chadha. 2005. Amplified fragment length polymorphism (AFLP) analysis of genetic diversity in Indian mungbean [*Vigna radiata* (L.) Wilczek] cultivars. Indian J. Biotech. 4:56-64.

Bhat, S.R., and B.S. Gill. 1985. The implication of 2n egg gametes in nobilisation and breeding of sugarcane. Euphytica 34:377–384.

Bremer, G. 1961. Problems in breeding and cytology of sugarcane. Euphytica 10:59–78.

Chiang, Y.S., and C.C. Lo. 1993. Genetic relationships among major sugarcane cultivars in Taiwan. Report of the Taiwan Sugar Res. Inst. 140:1-10.

Cox, T.S., Y.T. Kiang, M.B. Gorman, and D.M. Rodgers. 1985. Relationship between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci. 25:529-532.

D' Hont, A., L. Grivet, P. Feldman, P.S. Rao, N. Berding, and J.C. Glaszmann. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Mol. Gen. Genet. 250:405–413.

Deren, C.W. 1995. Genetic base of U.S. mainland sugarcane. Crop Sci. 35:1195-1199.

Efron B. 1982. The bootstrap, the jackknife, and other resampling plans. CBMS-NSF Regional conference series in Applied Mathematics No. 38. Society for Industrial and Applied Mathematics, Philadelphia.

Ferriol, M., B. Picó, P.F. de Córdova, and F. Nuez. 2004. Molecular diversity of a germplasm collection of squash (*Cucurbita moschata*) determined by SRAP and AFLP markers. Crop Sci. 44:653-664.

Graner, A., W.F. Ludwig, and A.E. Melchinger. 1994. Relationships among European barley germplasm: II. Comparison of RFLP and pedigree data. Crop Sci. 34:1199-1205.

Hammer, Ø., D.A.T. Harper, and P.D. Ryan. 2001. Past: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica, vol. 4, issue 1, art. 4: 9pp., 178kb. http://palaeo-electronica.org/2001_1/past/issue1_01.htm.

Hu, J.G., and B.A. Vick. 2003. Target Region Amplification Polymorphism: A novel marker technique for plant genotyping. Plant Mol. Biol. Rep. 21:289–294.

Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. Bulletin de la Société Vaundoise des Sciences Naturelles 44:223–270.

Kempthorne, O. 1957. An Introduction to Genetic Statistics. Wiley and Sons, New York.

Kim, H.S., and R.W. Ward. 1997. Genetic diversity in Eastern U.S. soft winter wheat (Triticum aestivum L. em. Thell.) based on RFLPs and coefficients of parentage. Theor. Appl. Genet. 94:472-479.

Li, G., and C.F. Quiros. 2001. Sequence related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: Its application to mapping and gene tagging in Brassica. Theor Appl Genet. 103:455–461.

Lima, M.L.A., A.A.F. Garcia, K.M. Oliveira, S. Matsuoka, H. Arizono, C.L. de Souza, Jr., and A.P. de Souza. 2002. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum* spp.). Theor. Appl. Genet. 104:30–38.

Ma, H.M., S. Schultze, S. Lee, M. Yang, E. Mirkov, J. Irvine, P. Moore, and A.H. Paterson. 2004. An EST survey of the sugarcane transcriptome. Theor. Appl. Genet. 108:851-863.

Mangelsdorf, A. 1983. Cytoplasmic diversity in relation to pests and pathogens. Sugarcane Breeders' Newsletter 45:45-49.

Mantel, N. 1967. Detection of disease clustering and generalized regression approach. Cancer Res. 27:209-220.

Powell, W., M. Morgante, J.J. Doyle, J.W. McNicol, S.V. Tingey and A.J. Rafalski. 1996. Genepool variation in genus *Glycine* subgenus *Soja* revealed by polymorphic nuclear and chloroplast microsatellites. Genetics. 144:793-803.

Rohlf, F.J. 2000. NTSYS-PC numerical taxonomy and multivariate analysis system ver 2.2. Applied Biostatistics, NY.

Rohlf, F.J., and R.R. Sokal. 1981. Comparing numerical taxonomic studies. Syst. Zoo. 30:459- 490.

Rozen, S., and H.J. Skaletsky. 2000. Primer 3 on the WWWfor general users and for biologist programmers. Pages: 365–386. In: Bioinformatics Methods and Protocols: Methods in Molecular Biology, S. Krawetz and S. Misener (ed.). Humana Press, Totowa, NJ.

SAS Institute Inc. 2002. SAS. Statistical Analysis Software for Windows, 9.0 ed. Cary, NC. USA.

Selvi, A., N.V. Nair, J.L. Noyer, N.K. Singh, N. Balasundaram, K.C. Bansal, K.R. Koundal, and T. Mohapatra. 2005. Genomic constitution and genetic relationship among the tropical and subtropical Indian sugarcane cultivars revealed by AFLP. Crop Sci. 45:1750-1757.

Sun, G., G. Wang-Pruski, M. Mayich, and H. Jong. 2003. RAPD and pedigree-based genetic diversity estimates in cultivated diploid potato hybrids. Theor. Appl. Genet. 107:110-115.

Tian, Q.Z., R.H. Zhou, and J.Z. Jia. 2005. Genetic diversity trends of common wheat (*Triticum aestivum* L.) in China revealed with AFLP markers. Genet. Resources and Crop Evolution. 52:325-331.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee, M. Hornes, A. Freijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: A new technique for DNA fingerprinting. Nucl. Acid Res. 23:4407–4414.

Vuylsteke, M., R. Mank, B. Brugmans, P. Stam, and M. Kuiper. 2000. Further characterization of AFLP data as a tool in genetic diversity assessments among maize (*Zea mays* L.) inbred lines. Mol. Breeding. 6:265–276.

Weir, B. 1990. Genetic data analysis: Methods for discrete population genetic data. Sinauer Assoc. Sunderland, MA.

CHAPTER 4 PREDICTING CROSS PROGENY PERFORMANCE IN SUGARCANE FROM COEFFICIENT OF PARENTAGE, AFLP AND TRAP DERIVED PARENTAL GENETIC RELATIONSHIPS

4.1 Introduction

Cultivated sugarcane (*Saccharum* species hybrids, $2n$ = 128-140) is a highly heterozygous and clonally propagated crop. The long duration of a sugarcane selection cycle (12-15 years) limits the rapid development of new cultivars. In addition, because sugarcane is a clonally propagated crop, no new opportunities exist for sexual recombination following crossing; and therefore, breeders have to capitalize on the initial genetic variation created during hybridization. Since, selection cannot salvage a good cultivar from a poor cross, the choice of parents is crucial as it predetermines the outcome of subsequent selections during the clonal selection stage.

A thorough knowledge of the genetic relationships that exists in the available parental genotypes is fundamental to the success of a sugarcane (or any) breeding program. Usually, in a typical sugarcane improvement program, a vast number of crosses are produced and a majority of them are discarded based on the *per se* performance of the progeny. This process involves substantial investments in time, labor and money. Therefore, the ability to predict the breeding potential of a cross based on the choice of parents would be of immense value to sugarcane breeders as it would help concentrate efforts on the most promising crosses or base populations.

The breeding prospects of a base population could be assessed by the 'usefulness' criterion (Schnell, 1983). 'Usefulness' is defined as the sum of base population mean and the selection intensity. According to Falconer and Mackay (1996), the selection intensity is a function of genetic variance (σ^2_g) and heritability of the trait with respect to a base population. Therefore, the 'usefulness' accounts for the differences in both means

(superior progeny *per se* mean i.e., heterosis) and genetic variances of the base populations. In an applied plant breeding program, base populations with high genetic variance are more desirable because transgressive segregants are more likely to be identified and subsequently selected. It was opined that heterosis and genetic variance are the functions of genetic divergence (or genetic distance, GD) between the parents (Moll et al., 1965; Bohn et al., 1999).

The relationship between GD among the parents and the genetic parameters (such as mean, heterosis and σ^2_g of progeny) have been studied in a wide array of crops using either coefficient of parentage (*f*) (Cox et al., 1985; Souza and Sorrells, 1991; Manjarrez-Sandoval et al., 1997) or molecular markers (RFLPs by Lee et al., 1989 and Smith et al., 1990; RAPDs by Santos et al., 1994 and Lanza et al., 1997; AFLPs by Burkhamer et al., 1998 and SSRs by Bohn et al., 1999). In most cases, heterosis of cross has been predicted accurately based on the GD between the parents. For example, a positive relationship has been observed between hybrid vigor and RAPD-GD in Ethiopian mustard (Tecklewood and Becker, 2006), hybrid vigor and SSR-GD in maize populations (Reif et al., 2003) and hybrid vigor and AFLP- and SSR-GD in Broccoli (Hale et al., 2007). Smith et al. (1990) observed that a combination of RFLP and pedigree knowledge could predict high yielding single-cross combinations in maize. On the other hand in the case of σ^2_g, the prediction was found to be inconsistent and remains unresolved (Melchinger, 1998).

In most of the crops, the molecular marker based GS estimates were not able to accurately predict the σ^2_g (Kisha et al., 1997; Manjarrez-Sandoval et al., 1997; Bohn et al., 1999). In oats, for example, the best predictor of σ^2_g was found to be *f* (Cowen and Frey, 1987) while RFLP-GS could not accurately predict the σ^2_g (Moser and Lee, 1994).

Souza and Sorrells (1991) also observed similar results with *f* but pointed out that the relationship was trait dependent. The unsatisfactory results from molecular markers to predict genetic variance could be attributed to the genome coverage of markers especially RFLPs, RAPDs and SSRs which yield a few polymorphic markers. Although this shortcoming could be overcome by using AFLP markers, which produce a vast number of polymorphisms, in most cases even the AFLP-GS could not accurately predict the σ^2_g. One proposition to accurately predict the genetic variance was to use markers tightly linked to quantitative trait loci (QTL) (Hayes et al 1997). Recently, sequence related amplification polymorphism (SRAP) markers (Li and Quiros, 2001) and target region amplification polymorphism (TRAP) markers (Hu and Vick, 2003) have been reported to target either exons or introns or gene rich regions across the genome. A positive relationship has also been observed between SRAP-GS and hybrid vigor in Broccoli (Hale et al., 2007). However, thus far, similar studies using TRAP markers have not been reported in any crop although the TRAP markers have been used to study the genetic relationships in sugarcane (Arro, 2005; Alwala et al., 2006).

Research associating GS with either family means or genetic variances are lacking in sugarcane. The objectives of this study were to estimate (*i*) genetic similarities (GS) among nine elite sugarcane cultivars using coefficient of parentage (*f*) method and the AFLP and TRAP marker techniques and (*ii*) to study relationship of family means and σ^2_g with various GS estimates in five crosses involving the nine parents.

4.2 Materials and Methods

4.2.1 Plant Material and Field Trial

The nine sugarcane parents used in the study are described in Table 2. 1 of Chapter 2. These genotypes are either experimental clones or cultivars adapted to

Louisiana's unique temperate climate (Table 4.1). TucCP77-42, a major cultivar in northern Argentina, was bred also in Houma, LA using Louisiana adapted clones as the recurrent parents. This group of genotypes serves as an important parental pool for sugarcane crossing in Louisiana.

Five crosses were produced involving the nine parents and thirty clones were randomly chosen from each cross for further field trials. The nine parents along with thirty clones per family were planted in the fall of 2003 at the Sugar Research Station farm, St. Gabriel, LA. Two stalks per clone were planted in single-row plots (3 meters long with a spacing of 1.8 meters between rows) in a randomized complete block design with two replications. Each replicate contained 30 sets of families with each set containing clones from all the five families. Standard Louisiana sugarcane cultural practices were implemented to grow the crop (Legendre, 2001).

Table 4. 1. Description of five crosses used in the study.

Family	Female parent	Male parent
A	HoCP92-624	HoCP91-552
B	LCP86-454	LCP85-384
C	HoCP95-951	HoCP96-540
D	TucCP77-42	LCP85-384
E	HO95-988	L99-238

Phenotypic data was recorded in late September to early October in the 2004 (plant cane) and 2006 (second ratoon) crops. Competent data were not recorded in 2005 as the plots were damaged by hurricanes Katrina and Rita. Stalk height (SH, cm) was measured as the length of the longest stalk (from base of the stalk to the first visible dewlap) and stalk diameter (SD, cm) was measured from the middle of the stalk with no reference to the bud groove. The number of millable stalks per plot were counted (SC) and estimated to number of stalks per hectare. Cane yield (CY, Mg ha^{-1}) was estimated

as the product of stalk number and weight of stalks (Mg) per hectare. Theoretical recoverable sugars (TRS) was estimated from pol and Brix values according to Gravois and Milligan (1992) assuming a fiber of 12.5%.

Data was subjected to analysis of variance (combined across two crop-cycles) using Proc GLM of SAS ver 9.1 (SAS Inc., Cary, NC) using the linear model

$$Y_{ijkl} = \mu + C_i + R_j + CR_{ij} + F_k + CF_{ik} + RF_{jk} + CRF_{ijk} + S_l(F_k) + CS_{il}(F_k) + RS_{jl}(F_k) + E_{ijkl}$$

where

Y_{ijkl} is the observation l in crop-cycle i, replication j and family k;

μ is the overall mean;

C_i is crop-cycle i;

R_j is the replication j;

CR_{ij} is replication j in crop-cycle i;

F_k is family k;

FC_{ik} is family k in crop-cycle i;

RF_{jk} is family k in replication j;

CRF_{ijk} family k in crop-cycle i, in replication j;

$S_l(F_k)$ is the clone l within family k;

$CS_{il}(F_k)$ is clone l within family k, in crop-cycle i;

$RS_{jl}(F_k)$ is clone l within family k, in replication j;

E_{ijkl} is the residual.

The family and clones within family were treated as fixed effects and the other sources of variation were treated as random effects. The mean squares were tested for significance using the appropriate error terms. Simple correlations were estimated among the traits using the Proc CORR procedure of SAS. The genetic variance (σ^2_g) for each

trait was calculated based on the family-wise analysis of variance estimates. Mid parent heterosis (MPH) for each clone within a family was estimated as $[(F_1\text{-MP})/MP] \times 100$ where MP $[=(P_1+P_2)/2]$ is the mid-parental value for a particular trait and the average MPH (AMPH) was estimated as ΣMPH/30. The number of positive heterotic clones within a family were counted and expressed as percentage (PHC).

4.2.2 DNA Extraction and Molecular Marker Assays

Young leaves were collected from each of the parental genotypes, frozen immediately and stored at -80 ^{o}C. Later, the leaves were ground in liquid nitrogen. Genomic DNA was extracted using the Plant DNeasy Mini Kit (Qiagen, Valencia, CA). DNA concentrations were estimated by known concentration of lambda DNA in 1 % agarose gel.

4.2.3 Estimation of Genetic Similarities

Three measures of genetic diversity estimation were employed: two molecular marker (AFLP and TRAP) techniques and one pedigree based method (Coefficient of parentage). The AFLP technique generates a vast number of polymorphic loci across the whole genome from a single selective PCR amplification whereas the TRAP technique generates polymorphisms from the gene rich regions of the genome as the forward (fixed) primer was designed from genes or EST sequences and the accompanied reverse (arbitrary) primer was designed to target introns or exons (Hu and Vick, 2003). The protocols for AFLP and TRAP technique have been described in Chapter 2. The coefficient of parentage (*f*) was calculated as previously described (Kempthorne, 1957; Chiang and Lo, 1993; Deren, 1995), but for proper comparison with AFLP- and TRAP-based GS estimates, the *f* values were multiplied by 2.

4.2.4 Estimation of Correlation Coefficients

Simple correlations were performed among various GS estimates and between GS and progeny mean, GS and genetic variance, GS and AMPH and GS and PHC using the Proc CORR option of SAS ver 9.1.

4.3 Results

4.3.1 Field Trial and Genetic Parameters

The results from combined analysis of variance (Table 4. 2) show that there were no significant differences ($P > 0.05$) for crop-year effects except for stalk diameter although the means were reduced from plant cane to second ratoon crop in all the traits. No significant differences ($P > 0.05$) were observed among the five families for all the five traits, however, significant differences ($P < 0.01$) were observed among the clones within families.

Table 4. 2. Mean squares from analysis of variance for stalk height (SH), stalk count (SC), stalk diameter (SD), cane yield (CY) and total recoverable sugars (TRS) of five crosses tested in two years 2004 and 2006.

Source of Variation	df	SH	SC	SD	CY	TRS
Crop	1	302739.26	2624.51	771.46*	2398.57	1020.65
Rep	1	4757.64	2487.80	8.05	190.85	313.88
Crop*Rep	1	5746.29**	340.26	3.96	33.91	2151.43*
Fam	4	945.57	138.92	23.22	17.20	238.40
Crop*Fam	4	973.11	18.41	3.67	8.92	22.88
Rep*Fam	4	866.22	219.86	15.45	19.49	367.03
Crop*Rep*Fam	4	339.48	128.60	5.22	10.78	301.65
Clone (Fam)	145	788.89**	252.39*	33.91**	23.51*	892.97**
Rep*Clone (Fam)	145	821.03**	259.09*	26.67**	25.04**	897.93**
Year*Clone (Fam)	145	460.65	151.91	13.76	12.77	348.82
Error	145	548.74	194.62	11.91	18.00	319.45
Total	599					

*, ** represent significant levels at 0.05 and 0.01 probabilities, respectively.

The correlation coefficients for SH were always significant and positive with the rest of the traits (Table 4. 3). Cane yield was also significantly correlated with all the traits except TRS. Positive correlations were observed for TRS with SH ($r = 0.96$**) and

with SC (r = 0.10*) whereas non-significant correlations were observed with SD (r = -0.02) and CY (r = 0.03). SD was negatively correlated with SC (r = -0.21*)

Table 4. 3. Simple correlation coefficients among five traits stalk height (SH), stalk count (SC), stalk diameter (SD), cane yield (CY) and total recoverable sugars (TRS) combined across two years 2004 and 2006.

	SC	SD	CY	TRS
SH	0.226**	0.206**	0.587**	0.962*
SC		- 0.214**	0.578**	0.102*
SD			0.249**	-0.023
CY				0.039

*, ** represent significant levels at 0.05 and 0.01 probabilities, respectively.

The means, genetic variances, average mid parental heterosis (AMPH), percent heterotic clones (PHC) and mid parental (MP) values within each family for all traits are presented in Table 4. 4. No significant differences were observed for means among all the families. Positive genetic variances were found in the family B for most of the traits followed by family D and E. The other families had negative variances probably due to small population size. The AMPH was found to be negative in all the families for the traits stalk height (SH) and cane yield (CY) whereas positive AMPH was observed for the stalk diameter (SD). Negative AMPH was also observed in all the families except family B for the trait stalk count (SC) and families A and E for theoretical recoverable sugar (TRS). With respect to PHC, in all the families positive heterotic progeny were observed for the traits SD and TRS and except in the family A for the traits SC and CY. For SH, only the families B and E had positive heterotic clones.

4.3.2 Variation for Molecular Markers

A total of 40 AFLP primer combinations (PC) were tested out of which 28 PC generated polymorphisms useful for screening the nine parents. A total of 1325 bands

Table 4. 4. Estimates of family means, standard errors (SE) of means, genetic variances (σ^2_g), average mid parent heterosis (AMPH), percent heterotic clones per family (PHC) and mid parental (MP) values for the five agronomic traits

Genetic parameters	Trait	A	B	C	D	E
Family mean†	SH	166.40a	165.39a	165.42a	171.08a	171.39a
SE		2.87	3.16	3.36	3.36	3.19
	SC	34.37a	34.37a	32.44a	34.83a	32.46a
		1.17	1.30	1.50	1.42	1.42
	SD	22.20a	21.74a	22.38a	23.05a	22.36a
		0.47	0.43	0.49	0.39	0.37
	CY	8.83a	8.64a	8.58a	9.60a	8.72a
		0.86	0.96	1.05	1.18	1.07
	TRS	230.31a	229.42a	228.20a	227.96a	232.04a
		2.20	2.30	2.27	2.59	2.00
σ^2_g	SH	-38.79	133.79	58.79	-81.38	-5.88
	SC	-31.52	5.37	-40.00	68.17	42.00
	SD	1.83	4.95	-1.07	-1.75	1.56
	CY	7.86	51.16	49.67	267.84	106.95
	TRS	-7.82	17.78	-63.00	25.29	-14.78
AMPH	SH	-19.37	-11.14	-16.92	-18.35	-12.36
	SC	-24.72	12.80	-14.71	-24.59	-12.38
	SD	10.83	7.60	9.78	17.70	10.41
	CY	-42.09	-10.46	-32.22	-38.09	-21.28
	TRS	0.58	-5.27	-11.93	-2.83	2.85
PHC	SH	0.00	16.66	0.00	0.00	3.33
	SC	0.00	70.00	20.00	10.00	30.00
	SD	66.66	63.33	76.66	93.33	46.66
	CY	0.00	23.33	3.33	3.33	16.66
	TRS	53.33	23.33	3.33	43.33	40.00
MP	SH	206.76	185.81	199.19	208.75	195.56
	SC	45.67	30.40	38.00	45.69	36.79
	SD	20.03	20.25	20.41	19.53	20.21
	CY	15.21	9.60	12.68	15.23	10.97
	TRS	229.34	242.29	256.17	232.18	238.73

† Estimates of means with same letter indicate that the means are non significant at 0.05 levels of LSD. SH, stalk height; SC, stalk count; SD, stalk diameter; CY, cane yield; TRS, theoretical recoverable sugars

were amplified out of which 686 (53%) were polymorphic bands (Table 4. 5). The average bands and average polymorphic bands per PC were found to be 47 and 24,

respectively. The average polymorphic information content (PIC) value across all the polymorphic bands was found to be 0.35.

All the 12 TRAP primer combinations produced polymorphic bands across the nine parents. A total of 444 bands were generated out of which 242 (55%) were polymorphic(Table 4. 5). The average bands and average polymorphic bands per PC were found to be 37 and 20, respectively. The average polymorphic information content (PIC) value was found to be 0.36.

Table 4. 5. Summary of markers produced from AFLP and TRAP techniques

	Primer Combinations	Total bands	Polymorphic bands	Total bands per PC†	Polymorphic bands per PC‡	PIC
AFLP	28	1325	686	47	24	0.35
TRAP	12	444	242	37	20	0.36

†PC, primer combination; ‡PIC, polymorphic information content

4.3.3 Estimates of Genetic Similarity (GS)

A summary of TRAP-GS, AFLP-GS and *f*-GS estimates for each cross is presented in Table 4. 6. The TRAP-GS estimates between pairs of genotypes ranged from 0.65 (HoCP92-624 x HoCP91-552 and) to 0.79 (LCP86-454 x LCP85-384) and the AFLP-GS estimates ranged from 0.73 (Ho95-988 x L99-238) to 0.77 (HoCP95-951 and HoCP96-540) indicating a high genetic similarity among the nine parental genotypes. The *f*-GS estimates among the nine genotypes varied from 0.10 (Ho95-988 x L99-238) to 0.36 (LCP86-454 x LCP85-384) indicating a low genetic similarity among the genotypes. Reasons for this disparity were explained in Chapter 3.

4.3.4 Correlations among Genetic Parameters

The correlation coefficients among various genetic parameters are presented in Table 4. 7. The correlations between family means and σ^2_g estimates were negative ($r \leq$ -0.68) for SH and SD whereas positive correlations were observed for the

Table 4. 6. The genetic similarity (GS) estimates between the parents of five crosses derived from AFLP, TRAP and coefficient of parentage (*f*)

Family	AFLP-GS	TRAP-GS	*f*-GS
A	0.74	0.65	0.20
B	0.74	0.79	0.36
C	0.77	0.76	0.30
D	0.75	0.78	0.14
E	0.73	0.72	0.10

Table 4. 7. Simple correlation coefficients among genetic parameters family means, genetic variances (σ^2_g), average mid parental heterosis (AMPH), percent heterotic clones per family (PHC) and mid parental (MP) values for the five traits under study.

Genetic parameters	Trait†	σ^2_g	AMPH	PHC	MP
Family Mean	SH	-0.68	0.01	-0.35	0.35
	SC	0.27	-0.03	-0.02	0.34
	SD	-0.90*	0.94**	0.64	-0.73
	CY	0.90*	-0.48	-0.38	0.64
	TRS	0.01	0.76	0.44	-0.29
σ^2_g	SH		0.70	0.79	-0.90*
	SC		-0.07	0.07	0.11
	SD		-0.73	-0.65	0.42
	CY		-0.17	-0.08	0.32
	TRS		0.49	0.61	-0.74
AMPH	SH			0.80	-0.93*
	SC			0.98**	-0.94**
	SD			0.70	-0.91*
	CY			0.98**	-0.97**
	TRS			0.89*	-0.81
PHC	SH				-0.88*
	SC				-0.94**
	SD				-0.62
	CY				-0.92*
	TRS				-0.97**

*, ** represent significant levels at 0.05 and 0.01 probabilities, respectively. †SH, stalk height; SC, stalk count; SD, stalk diameter; CY, cane yield; TRS, theoretical recoverable sugars

traits SC, CY and TRS (r = 0.01 to 0.90*). High positive correlations were observed between family means and AMPH for the traits SD and TRS ($r \geq 0.76$) and the rest of the

traits showed either negative or low correlations. Negative correlations were observed between the mean and PHC, for the traits SH, SC and CY (r = -0.02 to -0.38) whereas positive correlations were observed for SD and TRS. Positive correlations were observed between family means and MP for the traits SH, SC and CY ($r \geq 0.34$) whereas negative correlations were observed for SD and TRS ($r \leq -0.29$).

The σ^2_g component was highly correlated with AMPH for the traits SH and TRS ($0.49 \geq r \leq 0.70$) and negatively correlated for SC, SD and CY ($r \leq -0.07$). Between σ^2_g and PHC, positive correlations ($0.07 \geq r \leq 0.79$) were observed for all the traits except SD and CY ($r \leq -0.08$). Negative correlations were observed between σ^2_g and MP values for the traits SH and TRS ($r \leq -0.74$) and positive correlations were observed for the rest of the traits ($0.11 \geq r \leq 0.42$).

Average MPH was always positively correlated (in some cases significantly) with PHC (r = 0.70 to 0.98**) and always negatively correlated with MP values (r = -0.81 to -0.97**). Negative correlations were consistently observed between PHC and MP values (r = -0.62 to -0.97**) for all the traits.

4.3.5 Correlations between Genetic Parameters and GS

The estimates of correlation coefficients between various genetic parameters and GS estimated by the three methods for all the traits are presented in Table 4. 8. With the exception of mid-parental (MP) values, the correlation coefficients between AFLP-GS estimates and family mean, σ^2_g , AMPH and PHC were negative and non significant (except AMPH of TRS) in most of the traits.

In the case of TRAP, the exact opposite was noticed where negative correlations were observed between TRAP-GS and MP for all of the traits except TRS. With the rest of the genetic parameters, positive correlations were observed for most of the traits

Table 4. 8. Simple correlation coefficients between genetic parameters and genetic similarity (GS) estimates derived from AFLP, TRAP and coefficient of parentage (*f*) for the five traits under study

Genetic parameters	Trait†	AFLP-GS	TRAP-GS	*f*-GS
Family Mean	SH	-0.42	-0.02	-0.90*
	SC	-0.19	0.13	0.07
	SD	0.27	0.09	-0.67
	CY	0.00	0.21	-0.52
	TRS	-0.81	-0.61	-0.43
σ^2_g	SH	0.11	0.43	0.85
	SC	-0.42	0.36	-0.60
	SD	-0.49	-0.01	0.55
	CY	-0.04	0.46	-0.54
	TRS	-0.58	0.24	-0.17
AMPH	SH	-0.44	0.39	0.26
	SC	-0.19	0.52	0.68
	SD	0.12	0.12	-0.62
	CY	-0.34	0.49	0.42
	TRS	-0.90*	-0.57	-0.73
PHC	SH	-0.34	0.48	0.61
	SC	-0.23	0.59	0.59
	SD	0.65	0.42	0.08
	CY	-0.51	0.42	0.28
	TRS	-0.71	-0.57	-0.68
MP	SH	0.23	-0.40	-0.58
	SC	0.12	-0.47	-0.59
	SD	0.14	-0.09	0.54
	CY	0.26	-0.39	-0.44
	TRS	0.66	0.43	0.56

†SH, stalk height; SC, stalk count; SD, stalk diameter; CY, cane yield; TRS, theoretical recoverable sugars

barring some exceptions such as SH and TRS in family mean, SD in σ^2_g, TRS in AMPH and PHC.

In the case of coefficient of parentage (*f*), negative correlations were observed with mean (except SC), σ^2_g (except SH and SD) and MP values (except SD and TRS) for most of the traits. On the other hand, positive correlations were observed between *f*-GS and AMPH (except SD and TRS) and PHC (except TRS). The correlations between

combined average GS from all the three techniques and all the genetic parameters were more a representative of correlations between *f*-GS and the genetic parameters.

4.4 Discussion

4.4.1 Correlation between Molecular Marker-GS and *f* –GS Estimates

According to Melchinger et al. (1991), a correlation is expected between *f*-GS and marker-GS in related genotypes because of the linear relationship between the two measures. However, due to the basic differences in the assumptions/concepts underlying both measures deviations are expected as observed by Cox et al., (1985) in wheat (r = 0.27 between of *f* –GS and isozyme-GS) and Graner et al., 1994 in barley (r = 0.27 between *f* –GS and RFLP-GS). In this study, moderate (r = 0.42; refer to Chapter 3) correlation coefficients were observed between *f* –GS and marker-GS corroborating other previous reports in various crops using different molecular markers (Graner et al., 1994; Barbosa-Neto et al., 1996; Kim and Ward, 1997). The coefficient of parentage (*f*) determines the GD among genotypes assuming that the original ancestors are unrelated and the two alleles chosen at random from each parent are 'identical by descent' (Kempthorne, 1957).

In sugarcane, the assumptions of *f* may not hold true as very few parents were used in initial 'nobilization' and most of the parents for current cultivars are inherently related to each other. In addition, the effects of selection and genetic drift are completely ignored in the *f* method while in reality they introduce a bias and increase the standard error of *f*, thus further reducing the correlation between *f*-GS and marker-GS.

The molecular marker-GS estimates, on the contrary, are indicative of the proportion of bands shared by two genotypes and resemble DNA sequences across the whole genome (Nei, 1987). Therefore, the marker based GS estimates reflect the

proportion of 'genes alike in state' irrespective of whether they are identical by descent or only identical in state (Bohn et al., 1999). The marker-GS estimates are, however, estimated by assuming that all the co-migrating bands are identical in state but without sequencing it would be difficult to declare unequivocally that the bands or alleles that are supposedly identical in state are not co-migrating, non-homologous bands. A reliable and accurate estimate of marker-GS could be achieved by using a large number of molecular markers.

4.4.2 Relationship among the Genetic Parameters

In sugarcane cultivar development, Schnell's (1983) 'usefulness' concept is being adopted in a two stage selection process. Firstly, depending upon the significant mean differences among the families and available genetic variance, superior families are selected and secondly, depending upon the genetic variance among the progeny within a selected family, desirable selections are made among clones within a family to have substantial amounts of selection gain. The two genetic parameters (means and genetice variances) of Schnell's (1983) 'usefulness' model are discussed below

(i) Relationship between family means and other genetic parameters

According to Falconer and Mackay (1996), the selection gain for a highly heritable trait is predictable based on the indirect selection for family means. In sugarcane, Chang and Milligan (1992) observed that the family mean is a good indicator of the potential of a cross to produce elite progeny for a particular trait. Several studies in wheat and maize (Busch et al., 1974; Utz et al., 2001) also indicated that the family means were good indicators of progeny performance and could be accurately predicted by the mid parental (MP) values. However, Mather and Jinks (1982) stated that the family means prediction may not be accurate based on MP values if there is a prevalence

of either dominance or epistatic effects. In this study, the family means were not predicted accurately from the MP values and moderate correlations were observed between the family means and MP values which seemed to be trait dependent. These moderate correlations could be largely attributed to high degree of dominance or large epistatic effects which are more pronounced in a highly heterozygous polyploid crop like sugarcane. Previously, Utz et al., (2001) observed similar moderate correlations in wheat and in their study, epistasis seem to be the major cause of observed differences between family means and parental means.

Poor correlations between family means and σ^2_g for a particular trait are also not unusual because the mean values are influenced by masking effects of several genes and σ^2_g of a population is due to simultaneous segregation of several genes including those with masking effects. In this study, the correlations between family means and σ^2_g were either negative or low and when positive were non-significant. However, it is likely that the estimates of mean and σ^2_g may not reflect true values of a population as only thirty clones per cross were studied. Supporting our results were the observations from Kisha et al. (1997) (using a mean of top five yielding lines (M5TYL) in a population) and Utz et al. (2001) (using a mean of 22 individuals from a population) who could not draw reliable correlations between family means and σ^2_g. Hence in order to have accurate estimates of mean and σ^2_g with reduced standard errors, large number of clones per cross have to be evaluated which is not feasible from a practical point of view while handling a large number of crosses. Nevertheless, to achieve sizeable amounts of selection gains from a population, a combination of optimum levels of means and σ^2_g is necessary.

Interestingly, in this study, for TRS which is the economic trait of interest in sugarcane breeding, one promising observation was the relationship between family

means and AMPH and PHC. The family means were positively correlated with AMPH or PHC, indicating that the greater the number of heterotic clones per family, the higher the family mean. Although differences were not observed among family means in this study, in a breeding program where large number of clones per family are evaluated, the data suggests that heterotic clones could be selected from families with high means.

(ii) Relationship between σ^2_g and other genetic parameters

According to Gumber et al. (1999) a high correlation between σ^2_g and MP is expected only when all the QTL alleles are in coupling phase and have similar effect which might not hold true for sugarcane. They also stated that low correlations could also result (under extreme conditions) if the absolute difference between the parents ($|P_1-P_2|$) is zero (parents are highly genetically similar) and σ^2_g is high in a P_1 x P_2 cross when all the QTL segregating for the trait are in the heterozygous condition. In sugarcane since most of the QTL, if not all, are assumed to be in a heterozygous state, low correlations are to be expected between σ^2_g and MP values as seen in this study for the traits SC, SD and CY. In addition to the low correlations, negative correlations were also observed between σ^2_g and MP values for the traits SH and TRS.

Interestingly, for the traits SH and TRS, a positive relationship was observed between σ^2_g and AMPH and PHC. According to Melchinger et al. (1998), in autogamous crops a high correlation is expected between σ^2_g and MPH when the degree of dominance is always greater than zero (over dominance) at all the segregating QTL contributing to σ^2_g. Because sugarcane is highly heterozygous (or most of the segregating loci are highly heterozygous) a high degree of dominance is expected and consequently positive correlations were observed between σ^2_g and heterosis. However, Gumber et al. (1999) pointed out that a reduced correlation between σ^2_g and AMPH and PHC can be possible

in two cases: (*i*) even though a high degree of dominance is observed in F_1 hybrids, it is possible that most of the segregating QTL have only additive effects which mostly contribute to σ^2_g and not to MPH and (*ii*) if the parental alleles do not differ for their additive effects, some of the QTL with dominance effects would contribute more to MPH than to σ^2_g. In this study, although a high degree of dominance is expected in F_1 hybrids, it was not possible to study the QTL for its additive effects to verify the first case. On the other hand, for the second case, it is highly probable that the parental alleles might not differ for additive effects as high marker-GS were observed among the parental genotypes and thus might lead to moderate correlations between σ^2_g and AMPH and PHC for the traits SH and TRS. On the whole, it can be said that the moderate positive relationships between σ^2_g and heterosis (AMPH and PHC) are in accordance with the quantitative genetic theory as observed previously by Melchinger et al. (1998).

4.4.3 Predicting Genetic Parameters from GS Estimates

(i) AFLP based prediction

Negative correlations were previously observed between SSR-GS and panmictic MPH in maize (Reif et al., 2003), σ^2_g and RFLP-GS in soybean (Manjarrez-Sandoval et al., 1997; Kisha et al., 1997) and in oats (Moser and Lee, 1994) and σ^2_g and SSR-GS and AFLP-GS in maize (Bohn et al., 1999). One of the major reasons for these negative correlations is that in all the aforementioned molecular markers, the polymorphisms are generated from random portions of the genome with little knowledge on their linkage with QTL. Since a comprehensive knowledge of markers linked to QTL is still unknown, AFLP markers were employed to estimate GS under the assertion that they cover the whole genome uniformly. In this study, assuming an even genome coverage, 28 AFLP primer combinations were employed which generated 686/1325 polymorphic bands. Yet,

non significant (except for AMPH of TRS) and negative correlations were predominantly observed for most of the traits between AFLP-GS and all the genetic parameters (except MP values) indicating that the greater the similarity between the parents, the lower were the genetic parameter estimates.

It has been hypothesized that increasing the number and density of markers may not improve the association of GS and various predictors especially σ^2_g unless the markers are identical or tightly linked to QTL affecting a particular trait. In a study by Hayes et al. (1997), out of 322 AFLP markers, 12 markers which were highly discriminating among the barley genotypes were actually linked to QTL clusters. On the basis of their findings, it can be said that only those markers which are linked to QTL would accurately divulge the trait-based genetic diversity/similarity among genotypes and thus show promise in correctly predicting the correlations between GS and various genetic parameters as observed in sorghum (Jordan et al., 2003). Although the effects of mis-scorings and high standard errors among the GS estimates cannot be ruled out, from our results, the negative correlations between AFLP-GS and most genetic parameters indicate that crosses could be planned between the divergent parents to obtain superior progeny for most agronomic traits including TRS; however it remains to be seen if there is any linkage disequilibrium among the AFLP loci and genes affecting agronomic traits.

(*ii) TRAP based prediction*

With the development of modern bio-informatic tools and the availability of vast EST/gene sequences from various crop species, it may be possible to investigate Hayes et al. (1997) hypothesis. Making the investigation easy is the known fact that markers generated from one species are cross-transferable to other species within a family of similar taxons. Therefore, molecular markers could be developed potentially from within

genes affecting the traits from similar species and could be employed to study genetic diversity among genotypes. In this study, a relatively new molecular marker technique named target region amplification polymorphism (TRAP) was employed which amplifies polymorphisms from specified target regions of the genome. The primers used in this study were derived from EST/gene sequences of other grass species closely related to sugarcane.

Most of the relationships observed between TRAP-GS and genetic parameters were mostly exactly opposite to those observed between AFLP-GS and genetic parameters. Positive correlations were primarily observed between TRAP-GS estimates and AMPH, PHC, σ^2_g and family means for most of the traits (except TRS for family mean, AMPH and PHC). Other negative correlations noticed in these genetic parameters were infinitesimal. The TRAP-GS were based on only 242 polymorphic bands generated from only twelve TRAP primer combinations designed from EST/gene sequences specific for sucrose related genes. Therefore, the positive relationship between TRAP-GS and other genetic parameters for the traits SH, SD, SC and CY, cannot be wholly credited to the amplification nature of TRAP as the primer sequences were specific for sucrose related genes. However, it is unknown that the QTL affecting these traits has any linkage or linkage disequilibrium with sucrose related genes.

Specifically for the trait TRS, negative correlations were observed between TRAP-GS and AMPH or PHC indicating that the greater the similarity between parents at the sucrose loci, the lower the chance of producing heterotic clones. This could be a typical case, as observed in most of the crops, where higher GS among parents often results in low heterosis in progeny and breeders usually opt for divergent crosses. In sugarcane breeding, considering the trend of making crosses among the proven genotypes

for agronomic traits including sugar content, the high selection intensity for sucrose over the years might have resulted in accumulation of most loci affecting sucrose content. Therefore, since no new alleles exist affecting sucrose accumulation, high heterosis is not expected and introduction of new alleles from other untapped sources could be considered as an alternate option. In a preliminary study conducted at USDA-ARS, Houma, Louisiana (Anna Hale, unpublished results), high heterosis and high σ^2_g were observed for sucrose content when domesticated *S. officinarum* (high sucrose parent) and wild *S. spontaneum* (low sucrose parent) were crossed.

(iii) Coefficient of parentage (f) based prediction

Melchinger (1993) stated that within a set of closely related parents, the marker based GS were negatively correlated with heterosis. Therefore, the parents that are related by pedigree will have high marker based GS estimates and it is highly likely that progeny shows low estimates of heterosis. However, in this study, positive correlations were observed between *f*-GS and AMPH and PHC for most agronomic traits barring TRS, the economic trait of interest. These results were expected considering the trend in sugarcane breeding where a vast number of crosses are made among the phenotypically superior parental genotypes because the probability of recovering superior progeny is higher when both parents are themselves superior compared to when one parent is inferior as observed in Family B (derived from closely related parents LCP86-454 and LCP85-384).

In alfalfa, Kidwell et al. (1994) observed that the lack of linkage disequilibrium among the parental genotypes lead to low correlations between hybrid performance and genetic diversity (GD). In sugarcane, most of the current cultivars were developed from a narrow pool of parents. Sugarcane breeders routinely include vintage elite parents

(selected based on past performance) and potential parents (largely based on their performance in advanced stage trials) which might not have undergone more than 10 cycles of meiosis. Therefore, linkage disequilibrium is expected among the contemporary sugarcane parental genotypes and a positive correlation between *f*-GS and either AMPH and/pr PHC is not surprising as observed in this study.

However, one interesting aspect of this study is that the negative correlation between *f*-GS and genetic parameters for the trait TRS is similar to those observed between TRAP-GS and genetic parameters. This similarity again emphasizes that for the trait TRS, divergent crosses have to be planned to promote maximum genetic gains from the progeny.

4.5 Conclusions

The overall genetic similarities among a small group of nine sugarcane genotypes were high. This study indicated that family mean performance was a good indicator of the presence of elite progeny especially, for TRS. In addition, the positive correlations between σ^2_g and AMPH and/or PHC for the trait TRS indicates that progeny producing high sucrose levels could be selected from the range of genotypes present in elite crosses. For most agronomic traits, however, the means and σ^2_g could not be correlated accurately with heterosis.

By using two diverse types of molecular markers namely AFLP and TRAP, and pedigree based estimation of genetic similarity, we have drawn reliable information on genetic similarities/diversities among the genotypes. AFLP-GS seem to be a fair predictor of genetic parameters for most agronomic traits including TRS; however, it remains to be seen in this study if there exists any linkage disequilibrium between the marker loci and genes affecting the traits. As suggested by Hayes et al. (1997), QTL

linked AFLP markers could be a better option to accurately predict the genetic parameters in crops where saturated genetic linkage maps exist.

The idea behind using TRAP markers arose from suggestions from various studies that the accurate relationships can be derived between genetic parameters and marker-GS, if the markers are either linked to QTL regions or in linkage disequilibrium with genes. Indeed, for the trait TRS our results imply that there exists a relationship between genetic parameters and TRAP-GS which could however be functionally resolved either by sequencing the TRAP bands or by generating a TRAP based linkage map for QTL analysis. The predictions of genetic parameters derived from the *f*-GS were mostly similar to those obtained by TRAP-GS.

For all the three systems, there seem to be a negative correlation between GS and the genetic parameter estimates especially for TRS which signifies the option of a divergent cross between parents to maximize the family means, genetic variance and heterosis. However, as Gumber et al. (1999) stated that GS or GD estimates alone cannot explain the differences in means or variances and therefore adaptability should also be considered as it affects the gene expression for a particular trait in a particular environment. Souza and Sorrells (1991) proposed that breeders should concentrate more on the cross progeny derived from diverse and yet adapted parents so as to maximize the heterosis for agronomic traits

Overall, since, each technique revealed an independent but similar relationship between genetic parameters and GS for TRS, a combination of two or more techniques is recommended to accurately describe relationships between GS estimates and genetic parameters in sugarcane breeding programs.

4.6 References

Alwala, S., A. Suman, J. A. Arro, J. C. Veremis and C.A. Kimbeng. 2006. Target Region Amplification Polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. Crop Sci. 46:448–455.

Arro, J. A. 2005. Genetic diversity among sugarcane clones using Target Region Amplification Polymorphism (TRAP) markers and pedigree relationships. Masters Thesis submitted to Louisiana State University, Baton Rouge, LA.

Barbosa-Neto, J.F., M.E. Sorrells, and G. Cisar. 1996. Prediction of heterosis in wheat using coefficient of parentage and RFLP-based estimates of genetic relationship. Genome. 39:1142-1149.

Bohn, M., H. F. Utz and A. E. Melchinger. 1999. genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. Crop Sci. 39: 228-237

Burkhamer, R. L., S. P. Lanning, R. J. Martens, J. M. Martin and L. E. Talbert. 1998. Predicting progeny variance from parental divergence in hard red spring wheat. Crop Sci. 38: 234-248

Busch, R. H., J. C. Janke and R. C. Frohberg. 1974. Evaluation of crosses among high and low yielding parents of spring wheat (Triticum aestivum L.) and bulk prediction of line performance. Crop Sci. 14: 47-50

Chang, Y. S. and S. B. Milligan. 1992. Estimating the potential of sugarcane families to produce elite genotypes using univariate cross prediction methods. Theor. Appl. Genet. 84:662-671

Chiang, Y.S., and C.C. Lo. 1993. Genetic relationships among major sugarcane cultivars in Taiwan. Report of the Taiwan Sugar Res. Inst. 140:1-10.

Cowen, N. M. and K. J. Frey. 1987. Relationship between genealogical distance and breeding behaviour in oats (*Avena sativa* L.). Euphytica 36: 413-424

Cox, T. S., Y. T. Kiang, M. B. Gorman and D. M. Rogers. 1985. Relationship between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci. 25: 529-532

Deren, C.W. 1995. Genetic base of U.S. mainland sugarcane. Crop Sci. 35:1195-1199.

Falconer, D. S and T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, London.

Graner, A., W.F. Ludwig, and A.E. Melchinger. 1994. Relationships among European barley germplasm: II. Comparison of RFLP and pedigree data. Crop Sci. 34:1199-1205.

Gravois, K. A. and S. B. Milligan. 1992. Genetic relationships between fiber and sugarcane yield components. Crop Sci. 32: 62-66.

Gumber, R, K., B. Schill, W. Link, E. v. Kittlitz and A. E. MelchingerMean, genetic variance, and usefulness of selfing progenies from intra- and inter-pool crosses in faba beans (*Vicia faba* L.) and their prediction from parental parameters. 1999. Theor. Appl. Genet. 98: 569-580

Hale, A. L., C. A. Kimbeng, M. N. Nzaramba and M. W. Farnham. 2007. Heterosis for horticultural traits in Broccoli. Theor. Appl. Genet. (Accepted: In Press).

Hayes, P. M., J. Cereno, H. Witsenhoer, M. Kuiper, M. Zabeau, K. Sato, A. Kleinhofs, D. Kudrna, A. Kilian, M. Saghai-Maroof, D. Hoffman and the North American Barley Genome Mapping Project. 1997. Characterizing and exploiting genetic diversity and quantitative traits in barley (Hordeum vulgare) using AFLP markers. J. Quantitative Trait Loci 3(2).

Hu, J.G., and B.A. Vick. 2003. Target Region Amplification Polymorphism: A novel marker technique for plant genotyping. Plant Mol. Biol. Rep. 21:289–294.

Jordan, D. R., Y. Tao, I. D. Godwin, R. G. Henzell, M. Cooper and C. L. McIntyre. 2003. Prediction of hybrid performance in grain sorghum using RFLP markers. Theor. Appl. Genet. 106:559–567

Kempthorne, O. 1957. An Introduction to Genetic Statistics. Wiley and Sons, New York.

Kidwell, K. K., D. R. Woodfield, E. T. Bingham. and T. C. Osborn. 1994. Molecular marker diversity and yield of isogenic 2*x* and 4*x* single-crosses of alfalfa. Crop Sci. 34: 784-788.

Kim, H.S., and R.W. Ward. 1997. Genetic diversity in Eastern U.S. soft winter wheat (Triticum aestivum L. em. Thell.) based on RFLPs and coefficients of parentage. Theor. Appl. Genet. 94:472-479.

Kisha, T. J., C. H. Sneller and B. W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37: 1317-1325

Lanza, L. L. B., C. L. de Souza Jr., L. M. M. Ottoboni, M. L. C. Vieira and A. P. de Souza. 1997. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor. Appl. Genet. 94: 1023-1030

Lee, E. A., M. J. Ash and B. Good. 2007. Re-examining the relationship between degree of relatedness, genetic effects and heterosis in maize. Crop Sci. 47: 629-635.

Lee, M., E. B. Godshalk, K. R. Lamkey and W. W. Woodman. 1989. Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. Crop Sci. 29: 1067-1071

Legendre, B. L. (2001). Sugarcane production handbook. Louisiana State University Agricultural Center Publication # 2859.

Li, G., and C.F. Quiros. 2001. Sequence related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: Its application to mapping and gene tagging in Brassica. Theor Appl Genet. 103:455–461.

Mather, K and J. L. Jinks. 1982. Biometrical genetics. 3rd ed. Chapman and Hall, London.

Manjarrez-Sandoval, P., T. E. J. Carter, D. M. Webb and J. W. Burton. 1997. RFLP genetic similarity estimates and coefficient of parentage as genetic variance predictors for soybean yield. Crop Sci. 37: 698-703

Melchinger, A. E. 1993. Use of RFLP markers for analyses of genetic relationships among breeding materials and prediction of hybrid performance. In: Buxton, D. R et al. (ed) Int. Crop Sci. I., Crop Sci. Soc. Amer., Madison, WI. pp 621-628

Melchinger, A. E. 1998. Genetic diversity and heterosis. In: Proc. Int. Symp. The genetics and exploitation of heterosis in crops. Mexico City, August 17-22, 1997. CIMMYT, Mexico. (In press).

Melchinger, A. E., R. K. Gumber, R. B. Leipert and M. Vuylsteke. 1998. Prediction of testcross means and variances among F_3 progenies of F_1 crosses from testcross means and genetic distances of their parents in maize. Theor. Appl. Genet. 96: 503-512

Melchinger, A. E., M. M. Messmer, M. Lee, W. L. Woodman and K. R. Lamkey. 1991. Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci. 31: 669-678

Moll, R. H., J. H. Lonnquist, J. V. Fortuna and C. E. Johnson. 1965. The relationship of heterosis and genetic divergence in maize. Genetics 52: 139-144

Moser, H and M. Lee. 1994. RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variance in oats. Theor. Appl. Genet. 87:947-956

Nei, M. 1987. Molecular evolutionary genetics. Columbia Univ. Press, New York.

Reif, J. C., A. E. Melchinger, X. C. Xia, M. L. Warburton, D. A. Hoisington, S. K. Vasal, G. Srinivasan, M. Bohn and M. Frisch. 2003. Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. Crop Sci. 43: 1275-1282

Santos, J. B. D, J. Nienhuis, P. Skroch, J. Tivang and M. K. Slocum. 1994. Comparison of RFLP genetic markers in determining genetic similarity among Brassica oleracea L. genotypes. Theor. Appl. Genet. 87: 909-915

Schnell, F. W. 1983. Probleme der Elternwahl - ein Überblick. In: Ber Arbeitstagg 1983 Arbeitsgem Saatzuchtleiter. Gumpenstein, Austria, pp 1-11

Smith, O. S., J. S. C. Smith, S. L. Bowen, R. A. Tenborg and S. J. Wall. 1990. Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, grain yield, heterosis and RFLPs. Theor. Appl. Genet. 80: 833-840

Souza, E. and M. E. Sorrells. 1991. Prediction of progeny variation in oat from parental genetic relationships. Theor. Appl. Genet. 82: 233-241

Teklewold, A and H. C. Becker. 2006. Comparison of phenotypic and molecular distances to predict heterosis and F1 performance in Ethiopian mustard (*Brassica carinata* A. Braun). Theor. Appl. Genet. 112:752–759

Utz, H. F., M. Bohn and A. E. Melchinger. 2001. Predicting progeny means and variances of winter wheat crosses from phenotypic values of their parents. Crop Sci. 41: 1470-1478

CHAPTER 5 LINKAGE MAPPING AND GENOME ANALYSIS IN A *SACCHARUM* INTERSPECIFIC CROSS USING AFLP, SRAP AND TRAP MARKERS

5.1 Introduction

Sugarcane belongs to the genus *Saccharum* of the grass family Poaceae. Modern sugarcane cultivars (2*n*=100 -140) are highly heterozygous and complex aneuploids. They were derived from interspecific crosses between the domesticated species *S. officinarum* and it wild relative *S. spontaneum.* The initial interspecific hybrids were repeatedly backcrossed to *S. officinarum.* This process, termed as 'nobilization', retained the high sugar producing ability of *S. officinarum* and minimized the negative effects of *S. spontaneum* (Roach, 1972; Sreenivasan et al., 1987). The nobilization process also resulted in improved cane yields, ratooning ability and increased resistance to biotic and abiotic stresses. During nobilization the progeny inherited *2n* gametes from the *S. officinarum* parent (Bhat and Gill, 1985; Bremer, 1961). Consequently, the genome composition of current sugarcane cultivars is approximately 80 % *S. officinarum*, 15 % *S. spontaneum* with 5% recombinant chromosomes (D'Hont et al., 1996).

S. officinarum clones are characterized by their thick stalks, high sucrose and low fiber content. Chromosome numbers have consistently been reported as $2n = 80$ with $x = 10$ as the basic chromosome number (Sreenivasan et al., 1987; Daniels and Roach, 1987; D'Hont et al., 1998). *S. spontaneum* clones, on the other hand, are characterized by thin stalks, low sucrose, high fiber, profuse flowering, good ratooning ability, and high levels of disease and insect resistance. Chromosome numbers have been reported to range from $2n = 40$ to 128 (Panje and Babu, 1960) with $x = 8$ as the basic chromosome number (Al-Janabi et al., 1993; da Silva et al., 1993; D'Hont et al., 1998).

Limited progress has been made in improving sugar content in most sugarcane breeding programs especially in the last decade (Jackson, 2005). One reason that has been proposed for the lack of progress is the narrow genetic base of sugarcane cultivars. Most of the present day cultivars around the world can be traced back to a few progenitors used in the initial interspecific hybridizations during nobilization (Berding and Roach, 1987; Deren, 1995). In most sugarcane breeding programs, crosses are made among existing improved clones followed by clonal selections. Following the success of nobilization, only sporadic efforts were made by a few breeding programs to utilize new sources of wild germplasm for sugarcane improvement (Berding and Roach, 1987). There is renewed interest among sugarcane breeders to explore wild germplasm for novel sources of genes that could be useful in sugarcane breeding programs.

Molecular markers can play a pivotal role in tracking favorable alleles from wild species as well as ascertaining their introgression into the cultivated background (Edmé et al., 2006). Genetic linkage maps generated from molecular markers have facilitated gene tagging, map based cloning and QTL mapping in many crops. They have also been useful for studying genome architecture and evolution, especially in interspecific crosses (deVicente and Tanksley, 1993).

The earliest molecular genetic linkage maps of the progenitors of modern sugarcane were developed in *S. spontaneum* using RFLP (da Silva et al., 1993; Ming et al., 1998) and RAPD markers (Al-Janabi et al., 1993) and in S. *officinarum* using RAPD (Mudge et al., 1996), and RFLP markers (Ming et al., 1998). Recently, Edmé et al. (2006) developed a *S. spontaneum* and S. *officinarum* map using SSR markers. AFLP markers have so far been used to construct genetic linkage maps of commercial sugarcane (Hoarau et al., 2001; Aitken et al., 2005; Reffay et al., 2005).

Molecular markers such as RFLP, RAPD, AFLP and SSR are ideal for genetic fingerprinting and construction of linkage maps. However, they do not use prior gene sequence information and produce polymorphisms randomly across the genome. Two new PCR-based markers namely, SRAP (sequence related amplified polymorphism) and TRAP (target region amplification polymorphism) which amplify intragenic polymorphism have been reported. SRAP markers are arbitrarily designed with an AT- and GC-rich motif to anneal to intron and exons, respectively (Li and Quiros, 2001). Sequenced SRAP amplicons from *Brassica* (Li and Quiros, 2001) and *Cucurbita* (Ferriol et al., 2003) when used in BLAST searches revealed significant similarities to reported gene sequences found in Genebank databases. With TRAP markers, a forward or fixed primer is designed using gene/EST sequence information, whereas the reverse primer is similar to a SRAP primer (Hu and Vick, 2003). Using TRAP primers designed from resistance gene analogs, Miklas et al. (2006) reported that some of the polymorphisms produced on a preexisting common bean (*Phaseolus vulgaris* L.) mapping population mapped to the vicinity of resistance gene QTLs. In sugarcane, BLAST searches using sequenced TRAP amplicons from a *S. spontaneum* clone revealed high homology with known gene sequences from other grass species. Remarkably, the search also pulled up the *S. officinarum* GenBank accession from which the fixed TRAP primer was designed (Alwala et al., 2006a). SRAP markers have been integrated into genetic linkage maps of brassica (Li and Quiros, 2001) and TRAP markers have been integrated into maps of wheat (Liu et al., 2005) and common bean (Miklas et al., 2006). In sugarcane, SRAP and TRAP markers are being used to characterize parental and wild germplasm collections (Alwala et al., 2006a, 2006b; Suman and Kimbeng, 2007) but their potential for linkage and QTL mapping is yet to be ascertained.

The objective in this study was to construct molecular linkage maps of sugarcane using AFLP in conjunction with SRAP and TRAP markers. The maps were based on an F_1 cross between two progenitor species of modern sugarcane namely, *S. officinarum* 'Louisiana Striped' and *S. spontaneum* 'SES 147B'. These framework linkage maps would foster our understanding of genome architecture and organization in the two species and lay the ground work for subsequent QTL studies.

5.2 Material and Methods

5.2.1 Plant Material and DNA Extraction

The mapping population consisted of 100 clones derived from a *S. officinarum* (Louisiana Striped, 2*n*=80) X *S. spontaneum* (SES 147B, 2*n*=64) cross made at the Sugarcane Research Unit, USDA-ARS, Houma, Louisiana. Leaf tissue from F_1 seedlings was collected, immediately frozen and later ground to powder in liquid nitrogen. Genomic DNA was extracted using Plant DNeasy Mini Kit (Qiagen, Valencia, CA) following the manufacturer's protocol. The concentration of DNA was estimated using known concentration of Lambda DNA in 1% (w/v) agarose gel.

5.2.2 AFLP Protocol

Genomic DNA (50 ng) was digested with *EcoR* I (6bp cutter) and *Mse* I (4bp cutter) restriction enzymes. Following the protocol of Vos et al. (1995), the digested DNA was ligated to *EcoR* I and *Mse* I adapters. Pre-amplifications were done using *EcoR* I + A and *Mse* I + C primers followed by selective amplifications using two selective nucleotides. Following a similar protocol, the AFLP procedure was also carried out using *Pst* I (6 bp cutter) and *Mse* I restriction enzymes. *EcoR* I and *Pst* I are methylation insensitive and sensitive, respectively. The PCR was carried out in a reaction volume of 10 µL consisting of 1 µL of 10X reaction buffer, 1.5 µL of 25 mM MgCl2, 1 µL of 10

mM dNTPs, 1 µL of 1 µM of E-ANN (IR-Dye labeled) and 1 µL of 10 µM forward primer and 0.2 µL of 5U *Taq* polymerase (Promega, Madison, WI). The reactions were run on an *i*-cycler (BioRad Labs, Hercules, CA). The PCR conditions for selective amplifications were as follows: initial denaturing step at 94 ^{o}C for 3 min followed by initial 12 cycles at 94 ^{o}C for 30 s, 65 ^{o}C for 30 s (with 1 ^{o}C decrement every cycle) and 72 ^{o}C for 1 min, then followed by 23 cycles at 94 ^{o}C for 30 s, 56 ^{o}C for 30 s, and 72 ^{o}C for 1 min with a final extension step at 72 ^{o}C for 7 min. A total of 20 *EcoR* I-*Mse* I and 15 *Pst* I-*Mse* I selective amplification primer combinations were used.

5.2.3 SRAP Protocol

The sequences of the forward and reverse SRAP primers used in this study are given in Table 5. 1. The forward primers were unlabeled whereas the reverse primers were labeled with either IR-700 or IR-800 dyes. PCR amplifications were carried out in a 10 µL reaction volume containing 1.5 µL of 10X PCR buffer, 1.0 µL of 25 mM MgCl2, 1.0 µL each of 10 mM forward and IR-700 and IR-800 dye labeled reverse primers, 1.0 µL of 10 mM dNTPs (Promega, Madison, WI), 0.2 µL of 5U *Taq* polymerase (Promega, Madison, WI) and 1.0 µL of 50 to 80 ng genomic DNA. The conditions for PCR were as follows: an initial denaturing step was performed at 94 ^{o}C for 4 min followed by 5 cycles at 94 ^{o}C for 45 s, 35 ^{o}C for 45 s and 72 ^{o}C for 1 min, followed by 35 cycles at 94 ^{o}C for 45 s, 53 ^{o}C for 45 s, and 72 ^{o}C for 1 min with a final extension step at 72 ^{o}C for 7 min. All the PCR reactions were performed on an *i*-cycler (BioRad Labs, Hercules, CA). A total of 32 SRAP primer combinations were used.

5.2.4 TRAP Protocol

The TRAP, like SRAP, is also a two primer PCR marker technique. The design of

Table 5. 1. The primer sequences (5'→3') used in the SRAP PCR amplifications.

Primer sequences	Primer labeling †
Reverse primers	
GAC TGC GTA CGA ATT AAT	IR-700 dye
GAC TGC GTA CGA ATT TGC	IR-700 dye
GAC TGC GTA CGA ATT GAC	IR-800 dye
GAC TGC GTA CGA ATT TGA	IR-800 dye
GAC TGC GTA CGA ATT AAC	IR-700 dye
GAC TGC GTA CGA ATT GCA	IR-700 dye
GAC TGC GTA CGA ATT CAA	IR-800 dye
GAC TGC GTA CGA ATT CAC	IR-800 dye
CGT AGC GCG TCA ATT ATG	IR-700 dye
GGA ACC AAA CAC ATG AAG A	IR-800 dye
Forward Primers	
TGA GTC CAA ACC GGA TA	-
TGA GTC CTT TCC GGT AA	-
TGA GTC CTT TCC GGT CC	-
TGA GTC CAA ACC GGA CC	-
TGA GTC CAA ACC GGA AG	-
TGA GTC CTT TCC GGT TAA	-

† Only the reverse primers were labeled with either the IR-700 or IR-800 dye to enable detection of amplified bands on the LI-COR DNA analyzer.

Table 5. 2. The fixed/forward primer sequences (5'→ 3') used in the TRAP PCR amplifications.

Gene/EST	Fixed primer sequence	NCBI GenBank accession number†
Sucrose Synthase (SuSy)	GGAGGAGCTGAGTGTTTC	AF263384
Soluble Acid Invertase (SAI)	AGGACGAGACCACACTCT	AF062735
Calcium Dependent Protein Kinase (CDPK)	ACAGAACCACCAAAGGAG	CF572977

† The fixed primers were designed from gene/EST sequences obtained from the NCBI GeneBank. The reverse primers were similar to SRAP reverse primers listed in Table 5. 1.

fixed/forward primers used in this study was previously described in Alwala et al. (2006a). In brief, three forward primers were designed using the gene/EST sequences of sucrose synthase (SuSy), soluble acid invertase (SAI) and calcium dependent protein kinase (CDPK). The genes SuSy and SAI are associated with sucrose metabolism whereas CDPK is believed to be associated with cold tolerance. The forward primer sequences are listed in Table 5. 2. The reverse primers employed were the same as the labeled SRAP primers (Table 5. 1). PCR amplifications were performed as described by Alwala et al. (2006a). A total of 17 TRAP primer combinations were used.

5.2.5 Marker Scoring

The PCR amplified products were run on a LI-COR 4300 sequencer (LI-COR Inc., Lincoln, NE). The gels were saved onto a computer and scored manually and independently by two people. Ambiguous data that could not be resolved between the two scorers were discarded. A pseudo test cross strategy was followed to score the polymorphisms (Grattapaglia and Sederoff, 1994). The bands were scored for presence or absence when heterozygous in one parent, null in the other and segregating in the F_1 population. The bands were divided into two groups, as *S. officinarum* and *S. spontaneum* bands, based on their parental origin. The polymorphic bands were then tested for 1:1 (single dose, SD) and 3.3:1 (double dose, DD) segregation ratios using χ^2- analysis. [If *Saccharum* were a disomic polyploid, the double dose segregation would be 3:1. On the other hand, if it were a polysomic polyploid, the double dose ratio would be 7:2 for *Saccharum officinarum* (x=10) and 11:3 for *S. spontaneum* (x=8). To overcome this complexity, we used a segregation ratio of ≤3.3:1 (√3x3.6:1) as it gives equal χ^2 value for 3:1, 7:2 and 11:3 ratios (Mather, 1957)]. The markers which did not fit into either single or double dose ratios were treated as segregation distorted markers. In addition, bands

present in both parents and segregating in a 3:1 fashion (simple duplex) in the F_1 population were retained for consensus linkage map construction.

5.2.6 Linkage Map Construction

The mapping software JoinMap ver 3.0 (van Ooijen and Voorrips, 2001) was used for map construction. Two genetic linkage maps, one for *S. officinarum* and one for *S. spontaneum*, were constructed at a LOD score of > 5.0. The Kosambi mapping function was employed with a recombination fraction of 0.45 to form LG. First, the maps were constructed using SD and DD markers. Then, markers showing segregation distortion were included in the final map. None of the distorted markers altered the final order of markers on the linkage group.

5.2.7 Tests for Type of Ploidy

Two tests, detection of repulsion phase linkages and χ^2 segregation ratios, were used in an attempt to infer ploidy type (polysomic *vs* disomic polyploid) in the *S. officinarum* and *S. spontaneum* parents.

To detect repulsion phase linkages, SD markers were inverted and combined with the original set of SD markers. Linkage maps were re-constructed using the new set of doubled SD markers. Presence of repulsion phase linkage is indicated by co-localization of the original SD marker with its corresponding inverted marker. The χ^2 tests were performed using single dose and double dose markers to further confirm the ploidy behavior in the two parents as described by da Silva et al. (1993).

5.2.8 Estimation of Genome Size and Genome Coverage

For each parental species, the genome size was estimated based on Hulbert et al. (1987) and the genome coverage was estimated based on Bishop et al. (1983) methods.

5.3 Results

5.3.1 Comparison among marker systems

Combined across the two parental species, 35 AFLP primer combinations produced a total of 409 polymorphic bands out of which 318 (77%) were SD markers and 50 (17%) were DD and 41 (6%) deviated from Mendelian ratios (i. e., distorted markers). The polymorphic AFLP bands varied from 4 to 29 with an average of 11.68 bands per primer combination (Table 5. 3). Among the AFLP primers, 20 *EcoR* I-*Mse* I primer combinations produced a total of 249 polymorphic bands out of which 198 (79%) were SD, 13 (5%) were DD and 38 (16%) were distorted markers. The 15 *Pst* I - *Mse* I primer combinations produced a total of 160 polymorphic bands out of which 120 (75%) were SD, 37 (23%) were DD markers and 3 (2%) were distorted markers (Table 5. 4).

Table 5. 3. Summary statistics of AFLP, SRAP and TRAP polymorphic markers segregating in the single dose (1:1) and double dose (3.3:1) ratios, and that deviated from these ratios (distorted markers) in the mapping population.[†]

	Polymorphic markers	Single dose markers	Double dose markers	Distorted markers
		35 AFLP primer combinations		
Total	409	318(78)	50(12)	41(10)
Range	4-29	3-15	0-6	0-7
Average	11.68	9.05	1.42	1.17
		32 SRAP primer combinations		
Total	160	92(58)	21(13)	47(29)
Range	1-10	0-8	0-2	0-4
Average	5	2.87	0.65	1.46
		17 TRAP primer combinations		
Total	81	59(73)	5(6)	17(21)
Range	1-12	0-8	0-1	0-3
Average	4.76	3.47	0.29	1.00

[†] Values in parenthesis indicate percentages

Table 5. 4. Summary of AFLP, SRAP and TRAP polymorphic markers segregating in the single dose (1:1) and double dose (3.3:1) ratios, and that deviated from these ratios (distorted markers) in the two *Saccharum* parental species. †

Marker type	Polymorphic markers	Single dose markers	Double dose markers	Distorted markers
		S. officinarum		
AFLP	233	183 (78)	20 (9)	30 (13)
EcoR I-*Mse* I	155	122 (79)	5 (3)	28 (18)
Pst I-*Mse* I	78	61 (78)	15 (19)	2 (3)
SRAP	75	37 (49)	11 (15)	27 (36)
TRAP	36	27 (74)	2 (6)	7 (20)
Total	344	247 (72)	33 (9)	64 (19)
		S. spontaneum		
AFLP	176	135 (77)	30 (17)	11 (6)
EcoR I-*Mse* I	94	76 (81)	8 (9)	10 (10)
Pst I-*Mse* I	82	59 (72)	22 (27)	1 (1)
SRAP	85	54 (64)	10 (12)	21 (25)
TRAP	45	32 (71)	3 (7)	10 (22)
Total	306	221 (72)	43 (14)	42 (14)

† Values in parenthesis indicate percentages

Combined across the two parents, the 32 SRAP primer combinations produced a total of 160 polymorphic bands out of which 92 (58%) were SD, 21 (13%) were DD and 47 (29%) were distorted markers. The total number of polymorphic SRAP bands varied from 1 to 10 with an average of 5 bands per primer combination (Table 5. 3).

Likewise, the 17 TRAP primer combinations produced a total of 81 polymorphic bands out of which 59 (71%) were SD and 5 (7%) were DD markers. Seventeen (22%) markers deviated from Mendelian segregation ratios. The total number of polymorphic

TRAP bands ranged from 1 to 12 with an average of 4.76 bands per primer combination (Table 5. 3).

5.3.2 Marker Segregation

Combined across the AFLP, SRAP and TRAP techniques, a total of 344 markers were heterozygous in *S. officinarum* 'Louisiana Striped' of which 247 (72%) were SD and 33 (10%) were DD markers. Sixty four (18%) markers deviated from Mendelian segregation ratios. In *S. spontaneum* 'SES 147B', 306 markers were heterozygous of which 221 (72%) were SD, 43 (14%) were DD and 42 (14%) were distorted markers. In both parental species, segregation distortion was highest in the SRAP markers followed by the TRAP and lowest in the AFLP markers (Table 5. 4).

5.3.3 Map Construction

The female parent *S. officinarum* 'Louisiana Striped' map comprised of 146 linked markers spread over 49 linkage groups (LG) (Fig 5. 1). The cumulative genome length covered was found to be 1732 cM. The LG length varied from 13 to 108 cM with an average of 12 cM between any two adjacent markers. The number of loci forming LGs varied from 2-13 with the more dense LGs (L1, L3, L15) being formed almost exclusively by AFLP markers. A majority of the 146 linked markers were AFLP (74%) with only 20% of SRAP and 6% of TRAP markers. Among the 108 AFLP markers, 92 were *EcoR* I –*Mse* I and 16 were *Pst* I- *Mse* I based markers.

The male parent *S. spontaneum* 'SES147B' linkage map comprised of 121 linked markers spanning 45 LGs with a cumulative genome length of 1491 cM (Fig 5. 2). The length of the LGs varied from 2 to 85 cM with an average of 12 cM between any two adjacent markers. The number of loci forming LGs varied from 2-12 with the densest LGs (S1) being formed almost exclusively by AFLP markers. Of the 121 linked markers,

65% were AFLP, 25% were SRAP and 10% were TRAP markers. Among the AFLP markers, 55 were generated by the *EcoR* I –*Mse* I and 24 by the *Pst* I- *Mse* I primer combinations.

5.3.4 Ploidy Type

No repulsion phase linkages were detected in *S.* o*fficinarum* 'Louisiana Striped', suggesting that it could be an autopolyploid. However, the χ^2- test results could not confirm the autoploidy behavior as significant estimates were observed in both the autoploidy and alloploidy tests suggesting it is not a strict autopolyploid (Table 5. 5).

On the other hand, for the *S. spontaneum* 'SES 147B', the non detection of repulsion phase linkages and the non significant χ^2 estimates for autoploidy strongly suggest it to be an autopolyploid (Table 5. 5).

Table 5. 5. Results from χ^2 tests to detect the type of ploidy in *S. officinarum and S. spontaneum* parents used in the study.

Marker class	Observed	Expected	
		Autoploid	Alloploid
		S. officinarum	
Single dose markers	247	218 (0.78)	210 (0.75)
Double dose markers	33	62 (0.22)	70 (0.25)
Total	280	$P< 0.00001$*	$P< 0.000001$*
		S. spontaneum	
Single dose markers	221	209 (0.79)	198 (0.75)
Double dose markers	43	55 (0.21)	66 (0.25)
Total	264	$P< 0.068^{NS}$	$P< 0.001$*

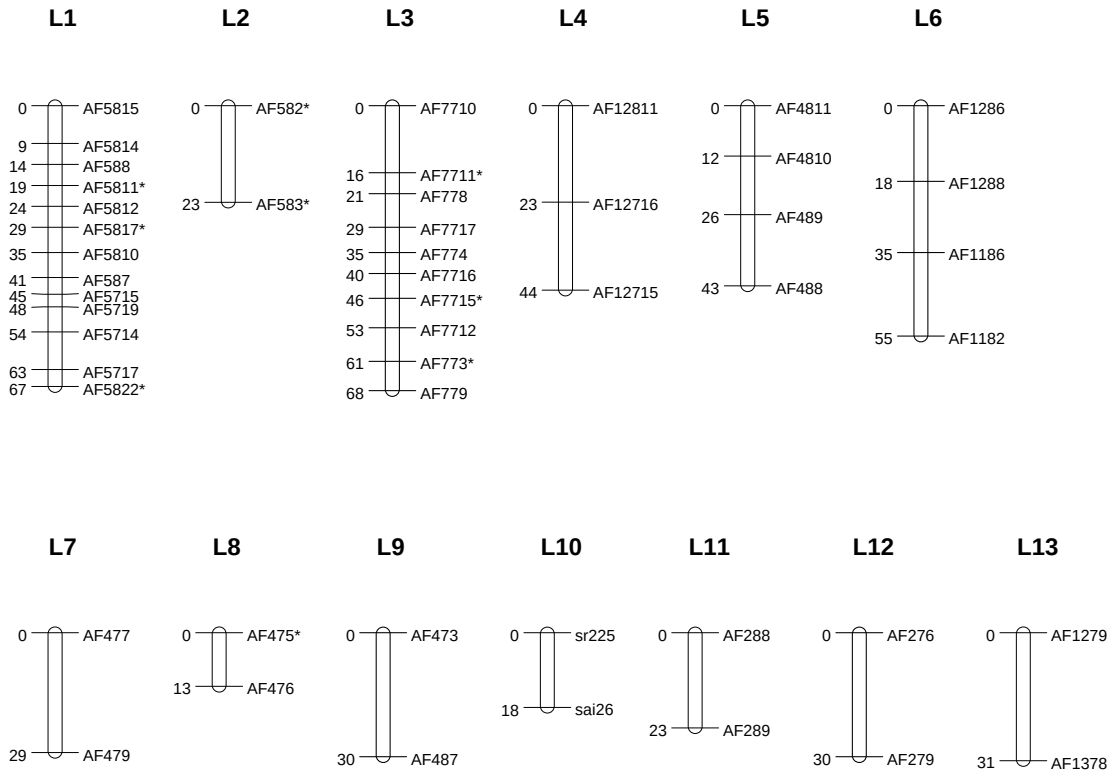
*, NS Indicates significance and non-significance at P = 0.05, respectively.

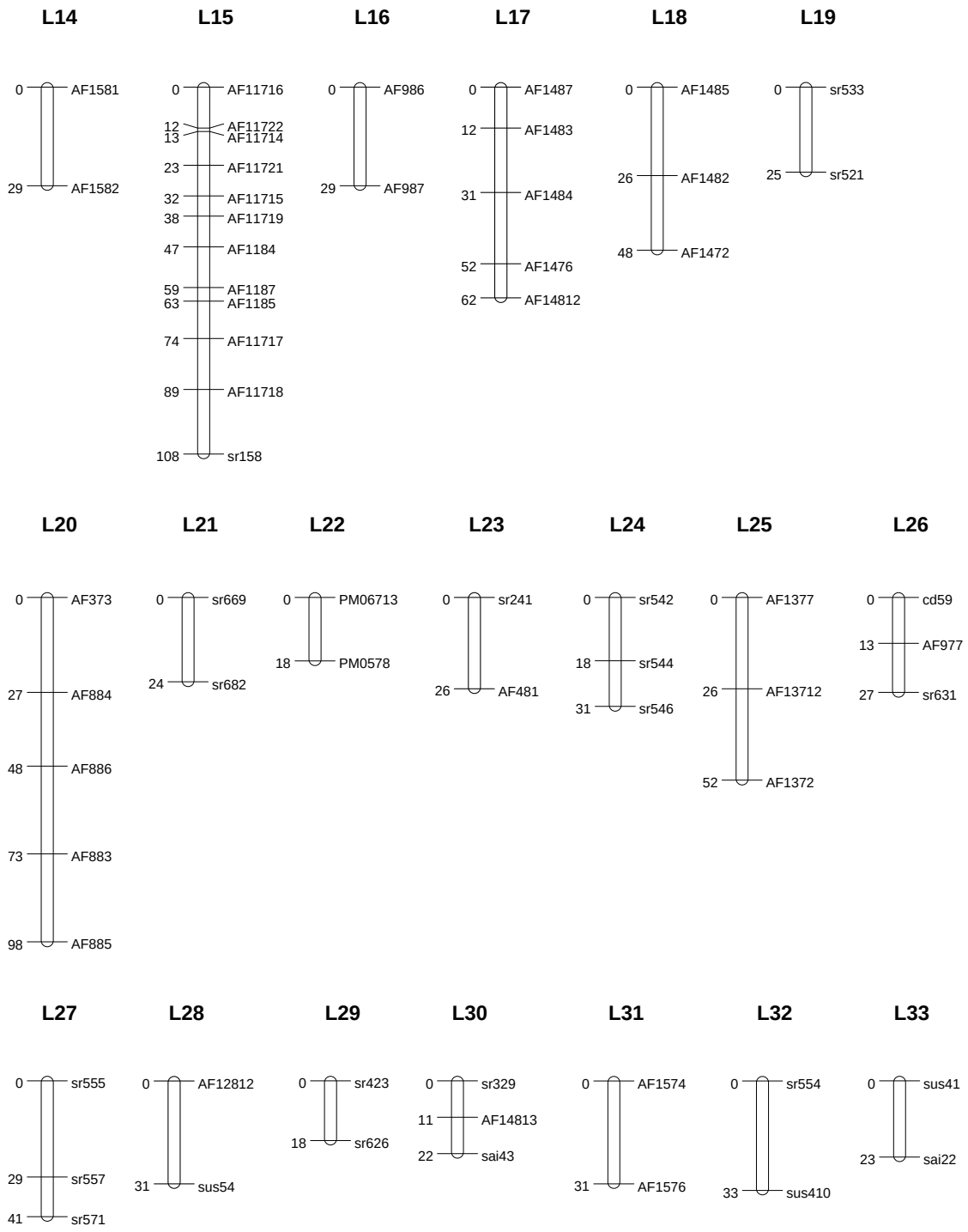
5.3.5 Genome Size and Genome Coverage

Three estimates of genome size were calculated at 10, 20 and 30 cM intervals

(Hulbert et al., 1987). For *S. officinarum*, using 32 (10 cM), 97 (20 cM) and 161 (30 cM) paired markers, the weighted genome size estimate was found to be 4897 cM . Likewise, for *S. spontaneum*, using 16, 53 and 90 paired markers, the weighted genome size estimate was found to be 6464 cM. However, as no repulsion phase linkages were observed, the estimates were divided by 2 (da Silva et al., 1993) giving rise to 2487 cM and 3232 cM genome size for *S. officinarum* and *S. spontaneum*, respectively. The computations indicated that approximately 69 % (1732 / 2487) of the *S. officinarum* genome and 46% (1491 / 3232) of the *S. spontaneum* has been covered.

From the Bishop et al. (1983) estimation, it was observed that there is 76% probability in *S. officinarum* and 63% in *S. spontaneum* to place a new marker (onto the constructed map) within a distance of 30 cM.




L14
0
AF1581
29
AF1582
L15
0
AF11716
12
AF11722
13
AF11714
23
AF11721
32
AF11715
38
AF11719
47
AF1184
59
AF1187
63
AF1185
74
AF11717
89
AF11718
108
sr158
L16
0
AF986
29
AF987
L17
0
AF1487
12
AF1483
31
AF1484
52
AF1476
62
AF14812
L18
0
AF1485
26
AF1482
48
AF1472
L19
0
sr533
25
sr521
L20
0
AF373
27
AF884
48
AF886
73
AF883
98
AF885
L21
0
sr669
24
sr682
L22
0
PM06713
18
PM0578
L23
0
sr241
26
AF481
L24
0
sr542
18
sr544
31
sr546
L25
0
AF1377
26
AF13712
52
AF1372
L26
0
cd59
13
AF977
27
sr631
L27
0
sr555
29
sr557
41
sr571
L28
0
AF12812
31
sus54
L29
0
sr423
18
sr626
L30
0
sr329
11
AF14813
22
sai43
L31
0
AF1574
31
AF1576
L32
0
sr554
33
sus410
L33
0
sus41
23
sai22

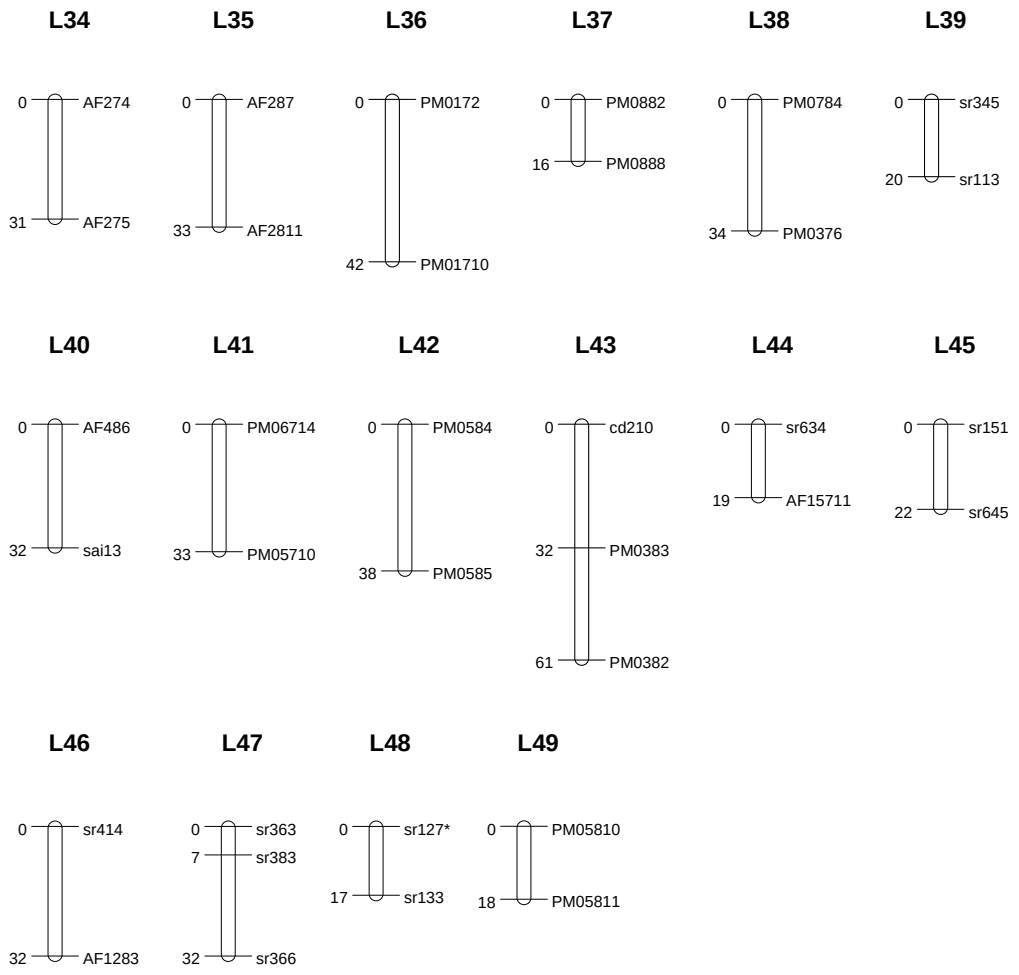


Figure 5. 1. Molecular marker linkage map of *S. officinarum* 'Louisiana Striped'. The map was constructed with a LOD score of > 5.0 and a recombination fraction of 0.45 using AFLP, SRAP and TRAP markers. Only single dose (1:1), double dose (3.3:1) and distorted markers were used to construct the linkage map. The Kosambi map distances (cM) and marker names are indicated on the left and right sides, respectively, of each linkage group. AFLP markers are denoted by 'AF' or 'PM', SRAP markers are denoted by 'sr' and the rest of the markers are TRAP markers. The numbers in each marker name represent the code used in our lab for primer combination along with the size of the band. The marker names with an asterix (*) represent distorted markers.

5.4 Discussion

5.4.1 Comparison among Markers

Two relatively new marker techniques, TRAP and SRAP, were used in conjunction with the AFLP marker technique for linkage mapping and analysis of an F_1 population derived from a cross between two progenitor species, *S. officinarum*

(Louisiana Striped, 2*n*=80) and *S. spontaneum* (SES 147B, 2*n*=64)), of modern sugarcane. The polymorphic fragments amplified by all three techniques were scored as dominant markers although Li and Quiros (2003) detected 20 % of codominant SRAP markers in a diploid brassica cross and Miklas et al. (2006) found 10 % of TRAP markers in a diploid common bean cross to be co-dominant. Due to the possibility of *2n + n* transmission, F_1 hybrids from a *S. officinarum* x *S. spontaneum* cross can harbor double the copy of homologous chromosomes from *S. officinarum* and variable copies of homeologous chromosomes from *S. spontaneum* (Bremer, 1961; D'Hont et al., 1996; Edmé et al., 2006). The difficulty of distinguishing multiple alleles from homologous and homoeologous chromosomes impedes the ability to determine homozygosity or heterozygosity and to designate co-dominant markers at any one locus. Therefore, methods for mapping in polyploid crops, such as sugarcane, have been developed which employ SD markers that are present in the simplex condition and segregate 1:1 irrespective of the ploidy level (Wu et al., 1992).

Single dose markers usually make up about 70 % of polymorphic loci detected in sugarcane mapping studies (Wu et al., 1992; da Silva et al., 1993; Hoarau et al., 2001; Aitken et al., 2005; Edmé et al., 2006; Garcia et al., 2006). These results are corroborated in this study by the frequency of SD markers reported for AFLP and TRAP but not for SRAP markers, which amplified only 58 % SD markers. Mapping in a complex polyploid, with large genome size such as sugarcane, requires substantially more markers and progeny than it would for a diploid. This makes SD markers the more important. Because of their relative abundance, SD markers facilitate mapping in polyploids by allowing the identification of alleles even in relatively small populations

(Wu et al., 1992). Deviations from the expected frequency of SD markers, especially in a small population, could seriously bias the outcome of linkage analysis. However,

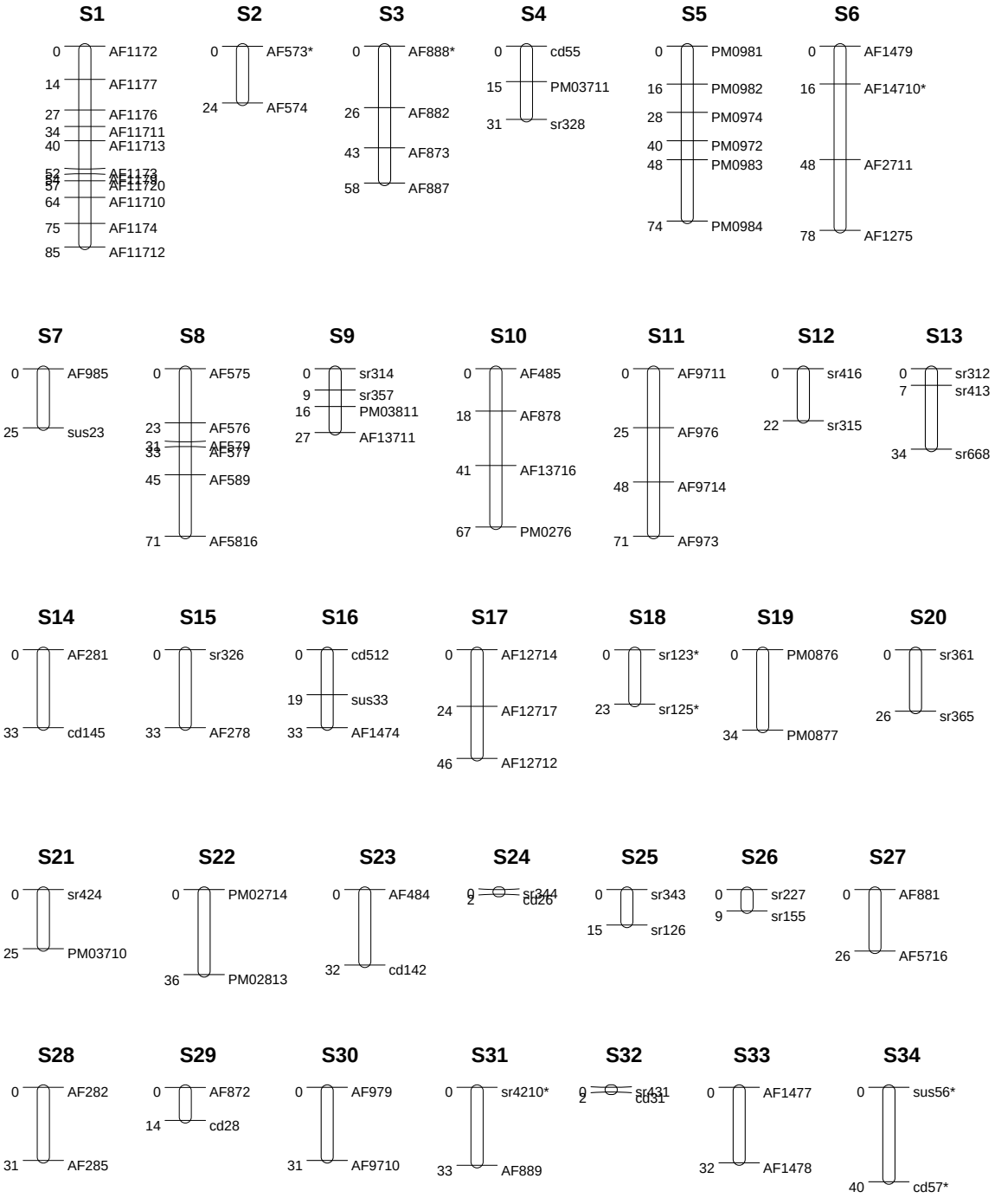


Figure 5. 2. Molecular marker linkage map of *S. spontaneum* 'SES 147B'. The map was constructed with a LOD score of > 0.5 and a recombination fraction of 0.45 using AFLP, SRAP and TRAP markers. Only single dose (1:1), double dose (3.3:1) and distorted markers were used to construct the linkage map. The Kosambi map distances (cM) and marker names are indicated on left and right sides, respectively of each linkage group. AFLP markers are denoted by 'AF' or 'PM', SRAP markers are denoted by 'sr' and the rest of the markers are TRAP markers. The numbers in each marker name represent the code used in our lab for primer combination along with size of the band. The marker names with an asterix (*) represent distorted markers.

as pointed out by da Silva et al. (1993), the proportion of SD polymorphism detected is not a function of the source of DNA used to detect the locus but rather that of the segregation at each locus according to the dosage of alleles at the locus. Therefore, it is less likely that SRAP markers possess an innate inability to amplify SD markers and rather coincidental that some of the SRAP primers used in this study did not align with SD loci in the population.

The TRAP and SRAP techniques target coding regions of the genome (Li and Quiros, 2003; Hu and Vick, 2003; Miklas et al., 2006) whereas, the AFLP technique amplifies a large number of random loci in a single assay (Vos et al., 1995). Therefore, as expected, the AFLP was by far the most superior marker technique for linkage mapping and analysis in a genome as large and complicated as that of sugarcane. The AFLP technique produced the most number of polymorphic bands and linked markers on

the map and the least number of distorted markers. Similar to reports in other crop species (van Heusden et al., 2000; Young et al., 2004; Mignouna et al., 2005a, b), the methylation sensitive *Pst* I – *Mse* I proved to be less robust compared to its methylation insensitive counterpart, *EcoR* I – *Mse* I, in generating data for linkage mapping. It has been suggested that DNA sequences are transcribed more readily from methylation sensitive regions of the genome (Cedar, 1988). The TRAP, SRAP and *Pst*-I – *Mse* I markers were included in this study for their potential to be associated with gene rich regions of the genome (Li and Quiros, 2003; Hu and Vick, 2003; Miklas et al., 2006; Cedar, 1988) as a prelude to our future goal of mapping QTLs associated with agronomic traits in these progenitor species. However, these markers have proved to be less efficient as tools for rapidly generating a large number of markers for linkage mapping, especially in genomes as large as that of sugarcane, because of the high throughput tools that would be needed to amplify numerous loci that are uniformly distributed across the genome. Miklas et al. (2006) and Li and Quiros (2003) utilized TRAP and SRAP markers, respectively, for QTL mapping by placing them onto pre-existing core maps. Most of the TRAP markers for disease resistance placed on the core map by Miklas et al. (2006) mapped to the exact same location. Liu et al. (2005) successfully used TRAP markers for linkage mapping in wheat only after lowering the initial annealing temperature. The authors conceded that the reverse primer, acting as a random primer, may have amplified fragments from various other regions in conjunction with itself. In preliminary trials, we found no tendency for TRAP primers to behave like RAPDs for the annealing temperature used in this study.

5.4.2 Map Construction

The pseudo-test cross strategy, based on an F_1 mapping population derived from crossing between two heterozygous parents, has been proposed for mapping outcrossing diploid and polyploid species for which inbred lines cannot be readily developed (Grattapaglia and Sederoff, 1994). This strategy has been widely used along with SD markers for mapping in polyploid species including sugarcane (Wu et al., 1992; da Silva et al., 1993; Mudge et al., 1996; Hoarau et al., 2001; Garcia et al., 2006; Edmé et al., 2006). Single dose markers, as earlier mentioned, can be detected in high frequencies even in relatively small populations. For example, irrespective of the ploidy level (i.e., $2n = 4X, 6X, 8X$ and $10X$), a population size of 75 individuals is considered large enough to detect SD loci at high confidence levels (Wu et al., 1992). The appropriate frequency (> 70%) of SD markers was detected in this study using a progeny size of 100 individuals. Appropriate levels of SD markers have been reported in other sugarcane mapping studies using a progeny sizes of 84 (Mudge et al., 1996), 88 (da Silva et al., 1993) and 100 (Guimaraes et al., 1999; Garcia et al., 2006) individuals.

The pseudo-test cross strategy allows two sets of SD markers, each set specific only to one parent, to be identified resulting in two parental maps (Grattapaglia and Sederrof, 1994; Maliepaard et al., 1998). Using this strategy, several female and male linkage maps have been constructed in *Saccharum* interspecific crosses (Ming et al., 1998; Edmé et al., 2006). Using RAPD markers, Mudge et al. (1996) published a map of *S.* o*fficinarum* with 51 LG spanning 1152 cM, that of Ming et al. (1998) using RFLPs included 72 LG spanning 2304 cM, whereas the map of Edmé et al. (2006) using SSRs had 25 LG covering 1180 cM. All these studies involved interspecific crosses in which *S.* o*fficinarum* was used as the female parent. Used as the male parent, the *S. spontaneum*

maps from a *S.* o*fficinarum* 'Green German' x *S. spontaneum* 'IND81-146'cross published by Ming et al. (1998) and Edmé et al. (2006) contained 69 LG spanning 1303 cM and 11 LG spanning 614 cM, respectively. Comparable number of LG and map length were found in the female (*S..officinarum* ' Louisiana Striped'; 49 LG with 1732 cM) and male (*S. spontaneum* 'SES 147B'; 45 LG with 1491 cM) maps in our study. In addition, to our knowledge, this is only the first attempt to use either or combination of AFLP, SRAP and TRAP markers for constructing linkage maps in these progenitor species of sugarcane.

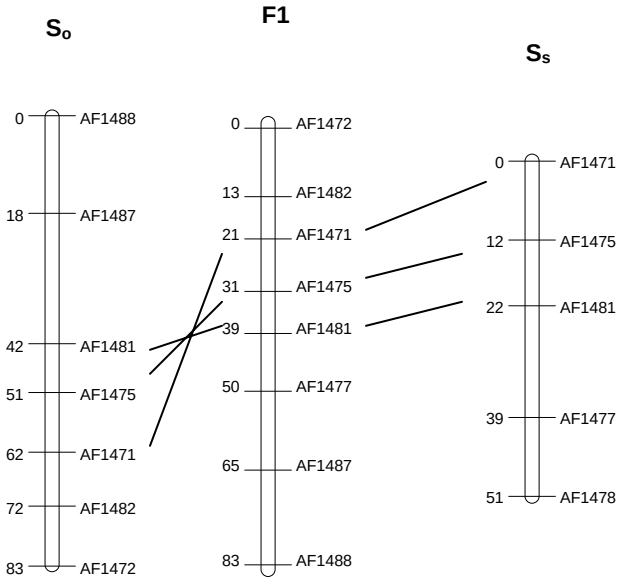


Figure 5. 3. An example of simple duplex markers mapping in *Saccharum* species. Simple duplex markers (common to both parents and segregate in 3:1 ratio in progeny) along with single dose (1:1), double dose (3.3:1) and distorted markers were used to construct the F_1 consensus linkage map and parental maps. S_o and S_s are linkage groups from the *Saccharum officinarum* and *S. spontaneum* maps, respectively, whereas F_1 is a linkage group from a consensus map of the two parents. The black lines represent the simple duplex markers common to both parents

The two framework maps in this study are unsaturated and cover only about 69 % and 46 % of the female and male genomes, respectively. This was quite evident from the

size of some LG (2 markers) and the substantial amount (60 %) of unlinked markers. Similar percentage of unlinked markers were reported by Garcia et al., (2005) in a sugarcane mapping population containing 100 individuals in contrast to the maps of Hoarau et al. (2001), Aitken et al. (2005) and Reffay et al. (2005) who reported about 15 % of unlinked markers using populations exceeding 200 individuals. The relatively small population size and stringent LOD score (> 5.0) used in our study were probably exacerbated by the complex genetic system inherent with interspecific crosses of sugarcane, leading to the high number of unlinked markers. However, although unsaturated, the stringent LOD score (> 5.0) employed and the relatively high proportion of SD compared to non-SD markers we used in building the maps provide two levels of confidence about the robustness of the maps, leaving little opportunity for spurious linkages. In fact, some of the small LG may actually be parts of larger groups which remained unconnected (Garcia et al., 2006). Our mapping population is being reconstituted to include more individuals and markers in an effort to saturate these framework maps for future use in QTL discovery.

5.4.3 Segregation Analyses

Although the two framework maps are far from saturated, marker segregation in the mapping population can be useful to unravel the genomic constitution and chromosomal behavior following hybridization of these two important progenitor species of modern sugarcane. For example, although not used for map construction, about 10 % of markers were simple duplex markers which are heterozygous in both parents and segregate 3:1 in the progeny. Although less informative for mapping, this class of markers may actually represent the degree of relatedness between the two mapping parents and could be useful as a locus bridge to form homology groups (Fig 5. 3)

(Grattapaglia and Sederoff, 1994; Malliepaard et al., 1998). This class of markers also portends the possibility that pairing and recombination can occur between chromosomes of the two species. In fact, *S. spontaneum* has been implicated in the ancestry of *S. officinarum* (Daniels *et al*., 1975) and *S. officinarum* genomes have been observed to contain *S. spontaneum* segments (Jannoo et al., 1998; D'Hont et al., 1989). Furthermore, using genomic *in situ* hybridization (GISH), Piperidis and D'Hont (2001) observed that between 5 to 17 % of the chromosomes in modern sugarcane cultivars had undergone recombination between the two progenitor species.

However, disparities exist between the two genomes and this can be inferred from the high proportion of distorted markers (18 % in *S. officinarum* and 14% in *S. spontaneum*) observed in this study since segregation distortion is a reflection of species relatedness or divergence (Tanksley and Nelson, 1996). Disparities exist between these two progenitor species in ploidy levels, chromosome numbers and genome size (Edmé et al., 2006; D'Hont et al., 1989). The possibility of $2n + n$ transmission in the progeny only serves as an added layer of complication. All of these factors can act independently or together to effect segregation distortion. In our study, the level of distortion was similar for both parents whereas using a similar type of population, Edmé et al. (2006) reported twice as much distortion (22%) with the female parent (*S. officinarum* 'Green German') and Ming et al. (1998) reported twice as much distortion (26%) with the male parent (*S. spontaneum* 'IND81-146').

Distorted markers may have biological significance if they are linked to lethal genes or loci with high genetic load causing inbreeding depression (Kuramoto et al., 1997; Barreneche et al., 1998). Maping may help unravel genomic regions which have high propensity for segregation distortion in sugarcane (Edmé et al., 2006). In pine

(Kubisiak et al., 1995) and oak (Barreneche et al., 1998) tree maps, distorted markers tended to cluster in particular LG. In sugarcane, Edmé et al. (2006) found evidence of clustering only for one linkage group and since the distorted markers were mostly in high dosage (non-SD), they postulated a possible role for double reduction in influencing distortion in that region of the genome. In this study, adding distorted markers onto the frame work maps, acquired one (L2) and two (S18 and S34) new LG in the *S. officinarum* and *S. spontaneum* maps, respectively, which were formed exclusively from distorted markers. Distorted markers also formed new LG with previously unlinked markers (L8, L 48, S2, S42, and S45) and a few of them mapped onto two preexisting LG (L1 and L3). However, non-biological factors such as small population size and fragment-complexes consisting of non-allelic co-migrating fragments can also effect segregation distortion. At least 30 % of the distorted AFLP alleles in a conifer cross was said to have resulted from fragment-complexes (Nikaido et al., 2000). More markers and possibly a larger population size would be necessary to accurately pin point, if present, loci harboring distorted markers and ascertain if such loci have biological significance in sugarcane.

The results from segregation and linkage analyses strongly suggest that the male parent *S. spontaneum* 'SES 147B' is an autopolyploid with chromosomes undergoing random pairing (polysomic segregation). This hypothesis has held true in all the studies attempted with *S. spontaneum* so far (da Silva et al., 1993; Al-Janabi et. al., 1993; Edmé et al., 2006). For the female parent, *S. officinarum* 'Louisiana Striped', linkage analysis failed to detect markers linked in repulsion phase but segregation analysis concluded that the genome was neither undergoing strict polysomic nor disomic segregation, thereby, concurring with previous evidence that *S. officinarum* could be an autoalloploid. Several authors (Mudge et al., 1996; Al-Janabi et al., 1994; Guimaraes et al., 1999; Edmé et al.,

2006) have found evidence of repulsion phase linkages in *S officinarum*, some using populations smaller than that used in this study.

5.5 Summary

Framework linkage maps of the two main progenitor species (*S. officinarum* and *S. spontaneum*) of modern sugarcane were constructed using AFLP with SRAP and TRAP markers. This is the first report using either one or a combination of AFLP, SRAP and TRAP markers to construct linkage maps in these species. SRAP and TRAP markers appeared less effective, compared to AFLP, as tools for rapidly generating a large number of markers for linkage mapping because of the high number of PCRs and high-throughput tools that would be required to amplify numerous loci genome-wide. However, SRAP and TRAP markers are attractive because of their potential for candidate gene analysis of QTLs, although that remains to be seen in sugarcane and is the subject of our subsequent study. Although the maps were incomplete, allele segregation in the mapping population allowed us to decipher genomic constitution and chromosomal behavior following hybridization of these two gene rich progenitor species upon which genetic improvement in sugarcane depends. Although the results revealed huge disparities in the two genomes, there seem to be sufficient similarity to support some level of pairing and recombination between chromosomes of the two species. Future studies are planned to include more individuals and markers to ensure better coverage of the genome in preparation for marker-assisted selection.

5.5 References:

Aitken, K. S., P. A. Jackson, and C. L. McIntyre. 2005. A combination of AFLP and SSR markers provide extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor. Appl. Genet. 110: 789-801

Al-Janabi, S. M., R. J. Honeycutt, M. McClelland and B. W. S. Sobral. 1993. A genetic linkage map pf Saccharum spontaneum L. 'SES 208'. Genetics 134: 1249-1260

Alwala, S., A. Suman, J. A. Arro, J. C. Veremis and C. A. Kimbeng. 2006a. Target Region Amplification Polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. Crop Sci. 46:448–455.

Alwala, S., C. A. Kimbeng, K. A. Gravois and K. P. Bischoff. 2006b. TRAP, a new tool for sugarcane breeding: comparison with AFLP and coefficient of parentage. Sugar Cane Intern. 24: 11-21

Barreneche, T., C. Bodenes, C. Lexer, J. F. Trontin, S. Fluch, R. Streiff, C. Plomion, G. Roussel, H. Steinkellner, K. Burg, J. M. Favre, J. Glössl and A. Kremer. 1998. A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 55 rDNA markers. Theor. Appl. Genet. 97: 1090-1103

Berding, N., and B. T. Roach. 1987. Germplasm, collection, maintenance and use. P. 143-210. In D. J. Heinz (ed.) Sugarcane improvement through breeding. Elsevier, Amsterdam

Bhat, S.R., and B. S. Gill. 1985. The implication of 2n egg gametes in nobilisation and breeding of sugarcane. Euphytica 34:377–384.

Bishop , D. T., C. Cannings, M. Skolnick and J. A. Williamson. 1983. The number of polymorphic clones required to map the human genome, pp. 118- 200 in Statistical Analysis of DNA Sequence Data, edited by B. S. WEIR. Marcel Dekker, New York.

Bremer, G. 1961. Problems in breeding and cytology of sugarcane. Euphytica 10:59–78.

Cedar, H. 1988. DNA methylation and gene activity. Cell 53:3-4

D' Hont, A., D. Ison, K. Alix, C. Roux and J. C. Glazmann. 1998. Determination of basic chromosome numbers in tha genus Saccharum by physical mapping of ribosomal RNA genes. Genome 41: 221-225

da Silva, J. A. G., M. E. Sorrells, W. Burnquist and S. D. Tanksley. 1993. RFLP linkage map and genome analysis of *Saccharum spontaneum.* Genome 36: 782–791.

Daniels, J., and B. T. Roach. 1987. Taxonomy and evolution in sugarcane. p. 7–84. In D.J. Heinz (ed.) Sugarcane improvement through breeding. Elsevier, Amsterdam.

Deren, C.W. 1995. Genetic base of U.S. mainland sugarcane. Crop Sci. 35:1195-1199

deVicente, M. C., and S. D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134: 585-596

D'Hont, A., L. Grivet, P. Feldmann, S. Rao, N. Berding and J. C. Glaszmann. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum spp*.) by molecular cytogenetics. Mol.Gen.Genet. 250: 405-413

Edmé´, S. J., N. G. Glynn and J. C. Comstock. 2006. Genetic segregation of microsatellite markers in *Saccharum officinarum* and *S. spontaneum.* Heredity 97: 366-375

Edmé´ S. J., J. C. Comstock, J. D. Miller and P. Y. P. Tai. 2005. Determination of DNA content and genome size in field grown sugarcane interspecific hybrids and genotypes. J Amer. Soc. Sugar Cane Technol. 5: 1–7

Ferriol, M., B. Picó and F. Nuez. 2003. Genetic diversity of a germplasm collection of Cucurbita pepo using SRAP and AFLP markers. Theor. Appl. Genet. 107: 271-282

Garcia, A. A. F., E. A. Kido, A. N. Meza, H. M. B. Souza, L. R. Pinto, M. M. Pastina, C. S. Leite, J. A. G. da Silva, E. C. Ulian, A. Figueira and A. P. Souza. 2005. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases Theor. Appl. Genet. 112: 298-314

Grattapaglia, D., and R. Sederrof. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137: 1121-1137

Guimarães, C. T., R. J. Honeycutt, G. R. Sills and B. W. S. Sobral. 1999. Genetic maps of *Saccharum officinarum* L. and *Saccharum robustum* Brandes & Jew. Ex Grassl. Genet. Mol.Biol. 22: 125-132

Guimarães, C.T., G. R. Sills and B. W. S. Sobral. 1997. Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proc. Natl. Acad. Sci. 94: 14261–14266.

Hoarau, J.Y., B. Offmann, A. D'Hont, A. M. Risterucci, D. Roques, J. C. Glaszmann and L. Grivet. 2001. Genetic dissection of a modern sugarcane cultivar (*Saccharum spp*.). I. Genome mapping with AFLP markers. Theor. Appl. Genet. 103:84–97

Hu, J.G., and B. A. Vick. 2003. Target region amplification polymorphism: A novel marker technique for plant genotyping. Plant Mol. Biol. Rep. 21:289–294.

Hulbert, S. H., T. W. Ilot, E. J. L. Egg, S. E. Lincolne, S. Lander and R. W. Michelmore. 1987.Genetic analysis of the fungus *Bremia lactucae* using restriction fragment length polymorphisms. Genetics 120: 947-958.

Jackson, P. A. 2005. Breeding for improved sugar content in sugarcane. Field Crops Research. 92: 277-290

Jannoo, N., L. Grivet, A. Dookun, A. D'Hont and J. C. Glazzmann. 1999. Linkage disequilibrium among modern sugarcane cultivars. Theor. Appl. Genet. 99: 1053-1060

Kubisiak, T. L., C. D. Nelson, W. L. Nance and M. Stine. 1995. RAPD linkage mapping in a longleaf pine x slash pine F_1 family. Theor. Appl. Genet. 90: 1119-1127.

Kuramoto, N., N. Tomaru, M. Murai and K. Ohba. 1997. Linkage analysis of isozyme and dwarf loci, and detection of lethal genes in sugi (*Cryptomeria japonica* D. Don). Breeding Science 47: 259-266

Li, G., and C. F. Quiros. 2001. Sequence related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: Its application to mapping and gene tagging in Brassica. Theor. Appl. Genet. 103:455–461.

Liu, Z. H., J. A. Anderson, J. Hu, T. L. Friesen, J. B. Rasmussen and J. D. Faris. 2005. A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci Theor. Appl. Genet. 111: 782-794

Maliepaard, C. , F. H. Alston, G. van Arkel, L. M. Brown, E. Chevreau, F. Dunemann, K. M. Evans, S. Gardiner, P. Guilford, A. W. van Heusden, J. Janse, F. Laurens, J. R. Lynn, A. G. Manganaris, A. P. M. den Nijs, N. Periam, E. Rikkerink, P. Roche, C. Ryder, S. Sansavini, H. Schmidt, S. Tartarini, J. J. Verhaegh, M. Vrielink-van Ginkel and G. J. King. 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor. Appl. Genet. 97: 60-73

Mather, K. (1957). The measurement of linkage in heredity. Methuen, London.

Mignouna, H. D. , R. A. Mank, T. H. N. Ellis, N. van den Bosch, R. Asiedu, S. Y. C. Ng, and J. Peleman. 2002a. A genetic linkage map of Guinea yam (*Dioscorea rotundata* Poir.) based on AFLP markers. Theor. Appl. Genet. 105: 716-725

Mignouna, H. D., R. A. Mank, T. H. N. Ellis, N. van den Bosch, R. Asiedu, M. Abang and J. Peleman. 2002b. A genetic linkage map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. Theor. Appl. Genet. 105: 726-735

Miklas, P. N., J. Hu, N. J. Grünwald, and K. M. Larsen. 2006. Potential application of TRAP (Targeted Region Amplified Polymorphism) markers for mapping and tagging disease resistance traits in common bean Crop Sci 2006 46: 910-916.

Ming, R., S. C. Liu, Y. R. Lin, J. A. G. da Silva, W. Wilson, D. Braga, A. van Deynze, T. F. Wenslaff, K. K. Wu, P. H. Moore, W. Burnquist, M. E. Sorrells, J. E. Irvine and A. H. Paterson. 1998. Detailed alignment of Saccharum and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Mudge, J., W. R. Andersen, R. L. Kehrer and D. J. Fairbanks. 1996. A RAPD genetic map of *Saccharum officinarum*. Crop Sci. 36:1362-1366

Nikaido, A. M., T. Ujino, H. Iwata, K. Yoshimura, H. Yoshimura, Y. Suyama, M. Murai, K. Nagasaka and Y. Tsumura. 2000. AFLP and CAPS linkage maps of *Cryptomeria japonica.* Theor. Appl. Genet. 100: 825-831

Panje, R. R., and C. N. Babu. 1960. Studies in *Saccharum spontaneum* distribution and geographical association of chromosome numbers. Cytologia 25: 152-172.

Reffay, N., P. A. Jackson, K. S. Aitken, J. Y. Hoarau, A. D'Hont, P. Besse and C. L. McIntyre. 2005. Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. Mol. Breed. 15: 367–381.

Roach, B.T. 1972. Nobilization of sugarcane. Proc. Int. Soc. Sugar Cane Technol. 14:206–216

Sreenivasan, T.V., B. S. Ahloowalia and D. J. Heinz. 1987. Cytogenetics. p. 211–253. In D.J. Heinz (ed.) Sugarcane improvement through breeding. Elsevier, Amsterdam.

Suman A and C. A. Kimbeng. 2007. Molecular diversity among ancestral species of *Saccharum* and the proportion of diversity represented in major Louisiana commercial breeding clones. ASA Southern Branch 2007 Annual Meeting February 4-6, 2007 Mobile, AL.

Tanksley, S. D., and J. C. Nelson. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92: 191-203

van Heusden, A. W. , J. W. van Ooijen, R. Vrielink-van Ginkel, W. H. J. Verbeek, W. A. Wietsma and C. Kik. 2000. A genetic map of an interspecific cross in *Allium* based on amplified fragment length polymorphism (AFLP™) markers. Theor. Appl. Genet. 100: 118-126

Van Ooijen, J. W., and R. E. Voorrips. 2001. JOINMAP 3.0. Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Freijters, J. Pot, J. Peleman, M. Kuiper and M. Zabeau. 1995. AFLP: A new technique for DNA fingerprinting. Nucl. Acid Res. 23:4407–4414.

Wu, K.K., W. Burnquist, M. E. Sorrells, T. L. Tew, P. H. Moore and S. D. Tanksley. 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. 83: 294–300.

Young, W. P., J. M. Schupp and P. Keim. 1999. DNA methylation and AFLP marker distribution in the soybean genome. Theor. Appl. Genet. 99: 785-792

CHAPTER 6 QUANTITATIVE TRAIT LOCI (QTL) ANALYSIS FOR SUGAR RELATED TRAITS IN A *SACCHARUM* INTERSPECIFIC CROSSS

6.1 Introduction

Cultivated sugarcane belongs to the genus *Saccharum* of the family Poaceae. The genus is clonally propagated and characterized by complex aneu-polyploidy and high heterozygosity. Modern sugarcane cultivars (*Saccharum* spp. hybrids, 2*n*= 100-140) are essentially the derivatives of interspecific crosses involving *S. officinarum* and *S. spontaneum* (although minor contributions from *S. barberi* and *S. sinense* were also observed). The interspecific hybridizations were initiated principally to transfer disease resistance genes from the wild *S. spontaneum* to the erstwhile domesticated *S. officinarum*. The interspecific hybrids were subjected to a series of backcrosses to *S. officinarum* to recover its high sucrose producing ability. This process termed as 'nobilization' provided a major breakthrough in the world's sugar industry via improved yields, increased productivity and better adaptability and ratooning ability (Sreenivasn et al., 1987).

"Nobilization" in sugarcane is perhaps the best example of the amazing contribution of wild germplasm to genetic improvement of an economically important crop (Martin, 1996). Unfortunately, very few parents were used in the initial hybridization event that occurred a century ago and, the derivatives from the event quickly became the genetic base for sugarcane breeding programs all over the world. In breeding sugarcane, crosses are made among the existing improved clones followed by clonal selections which typically last for about 10 - 15 years. Consequently, current sugarcane cultivars have undergone only about 10 cycles of recombination since nobilization. The genetic base of modern sugarcane is thus very narrow and this has been mentioned as one of the factors responsible for the slow progress currently being

experienced by some sugarcane improvement programs (Jackson, 2005). Mindful of the narrow genetic base of cultivated sugarcane and of the tremendous gains that could be achieved by tapping into novel alleles from the relatives of cultivated sugarcane, a basic breeding program was established by the USDA-ARS at Houma, Louisiana. Two main objectives of the program were: to broaden the genetic base of sugarcane and to identify and introgress useful genes from these wild relatives into the cultivated background. The release in 1996 of LCP 85-384 a widely adopted (> 85%) cultivar to the Louisiana industry is testament to the success of this program. This cultivar increased cane yield by 20-25% and contributed to substantial boosts in sugar production that was unprecedented in recent memory.

The basic breeding program was established in 1964 however, the first significant sugarcane cultivar from this program was not released until 30 years later. Efforts to identify and introgress novel genes from these wild relatives of sugarcane into the cultivated background can now be expedited through the use of molecular markers. In other crop species including tomato (Tanksley et al., 1996; Lecomte et al., 2004; Bernacchi et al., 2004) and corn (Bouchez et al., 2002), it has been possible using markers to detect quantitative trait loci (QTL) that control the genetic variability of complex traits and to introgress more than one QTL at a time through marker-assisted selection. Generally, marker-assisted introgression is expected to permit a gain time of about two to three backcross generations, compared to conventional backcross programs (Visscher et al., 1996).

Sugarcane has lagged behind other crops in utilizing molecular markers because of its complex genetics, i.e., large genome size, multiple alleles per locus from both homologous and homeologous chromosomes and the lack of diploid relatives with

simpler genomes that could be exploited to unravel the genetics of sugarcane. However, genetic tools for sugarcane are now becoming adequate to quantify the effects of genomic regions for traits displaying simple (Daugrois et al., 1996; Asnaghi et al., 2004; Raboin et al., 2006) as well as complex inheritance (Ming et al., 2001; Hoarau et al., 2002; Aitken et al., 2006; Al-Janabi et al., 2007) patterns.

Sugar yield is the most important trait in sugarcane improvement programs. Sugar yield is the function of cane yield and recoverable sucrose content from the harvested cane. Maximizing the sucrose content component of sugar yield would minimize harvesting, transporting and milling costs and lead to lucrative returns (Jackson and Morgan, 2003; Aitken et al., 2005). Sucrose content at the beginning of the crushing season (September in Louisiana) is typically low and increases as the cane matures. The deployment of cultivars high in sucrose at the commencement of the crushing season is one strategy to increase profitability in the Louisiana sugar industry. The 9-month growing season imposed by freezing temperatures in December, compared to 12-18 months in more tropical environments, further compels the necessity for early maturing cane in Louisiana. Research has shown that, for sucrose content, genetic variation and predicted gains from selection is typically higher in the earlier compared to the later months of the growing season (Cox et al., 1990) and that, selection practiced on the immature crop (early season) would have minimal adverse effects on the relative rankings of the mature crop (Jackson and Morgan, 2003). Breaux (1987) and Cox et al. (1994) further demonstrated that it was possible to breed high early sucrose sugarcane varieties using recurrent selection.

Whereas *S. officinarum* is unarguably the best source of genes for sugar related traits, recent evidence using molecular markers suggest that wild relatives such as *S.*

spontaneum (with relatively low sucrose content) cannot be discounted as potential contributors of novel genes for such a trait (Tanksley and McCouch, 1997). Reffay et al. (2005) found several markers associated with positive effects for sucrose related traits that were inherited from *S. spontaneum* whereas Ming et al. (2001) found markers associated with negative effects that were inherited from *S. officinarum*. Since both progenitor species can harbor alleles with both positive and negative effects, when selecting for say one major QTL, some knowledge about chromosomal segments or alleles outside the target QTL which may adversely impact the trait of interest would hasten selection during introgression and increase selection efficiency.

This study was undertaken to characterize QTLs controlling sugar related traits in these two progenitor species of modern sugarcane. The study was based on an F_1 cross between *S. officinarum* 'Louisiana Striped' x *S. spontaneum* 'SES 147B'. The traits were measured in the early and late plant growing seasons and over two crop-years. The objective was to detect QTLs controlling these traits and in particular to determine if different QTLs control these traits in the early and late part of the plant growing season and if QTLs stable across crop-years could be identified. In addition, the potential of discriminant analysis (DA) to identify molecular markers associated with sugar related traits was also evaluated by comparing the results from DA with those identified in QTL analysis. DA is a multivariate approach to genetic analysis which associates an individual with a descriptive class (Fisher, 1936).

6.2 Material and Methods

6.2.1 Plant Material and DNA Extraction

The mapping population consisted of 100 F_1 individuals derived from a cross between *S. officinarum* (Louisiana Striped, 2*n*=80) x *S. spontaneum* (SES 147B, 2*n*=64).

Leaf tissue from seedlings was collected, immediately frozen and later ground to powder in liquid nitrogen. Genomic DNA was extracted using Plant DNeasy Mini Kit (Qiagen, Valencia, CA) following the manufacturer's protocol. The concentration of DNA was estimated using known concentration of Lambda DNA in 1% (w/v) agarose gel. PCR protocols for AFLP, SRAP and TRAP where described elsewhere (Chapter 2).

6.2.2 Field Trial and Phenotyping

The F_1 progeny, the *S. officinarum* parent and four checks were grown at the USDA Sugarcane Research Unit farm, Chacahoula, Louisiana. The *S. spontaneum* parent could not be planted in the field as the USDA-APHIS rules do not permit the planting of weedy species in the field. The trial was planted in November 2003 in single-row 3 by 1.8 square meter plots in a randomized complete block design (RCBD) with three replications. Standard sugarcane cultural practices for Louisiana were followed (Legendre, 2001). In 2004 and 2005, ten stalks from each plot were randomly chosen and the juice extracted and used to estimate Brix (total soluble solids including sucrose in the juice sample), pol (estimate of sucrose in the juice sample) and sucrose content. The samples were collected during the early (late August) and late (just prior to harvest in late October) plant growing season. Brix was recorded in the laboratory using a RFM refractometer (Bellingham and Stanley$^+$ Co, England). Pol readings of the clarified juice were obtained using an automated saccharimeter (Kernchen, Germany). Theoretical recoverable sucrose content (hereafter refered as sucrose in this chapter) was calculated from the Brix and pol data assuming 12.5% of fiber in all the clones (Gravois and Milligan, 1992). In the 2004 plant cane crop only early season Brix (04EB) was measured whereas the late season data included Brix (04LB), pol (04LP) and sucrose (04LS). In the 2005 first ratoon crop the plots were inaccessible following the

devastation from hurricanes Katrina and Rita, thus, data were available only from the early season for Brix (05EB), pol (05EP) and sucrose (05 ES).

6.2.3 Genotyping and Linkage Map Construction

Amplified fragment length polymorphism (AFLP), sequence related amplified polymorphism (SRAP) and target region amplification polymorphism (TRAP) marker techniques were employed to generate polymorphic bands. Two genetic linkage maps were constructed: one for *Saccharum officinarum* 'Louisiana Striped' and the other for *S. spontaneum* 'SES 147B'. A total of 344 markers were generated from *S. officinarum* and a total of 306 markers from *S. spontaneum*. The framework *S. officinarum* map comprised of 146 linked markers spread over 49 linkage groups while the *S. spontaneum* map comprised of 121 linked markers in 45 linkage groups. A detailed mapping study is reported in Chapter 5.

6.2.4 Statistical Analysis

The phenotypic data was tested for normality using PROC UNIVARIATE option of SAS ver 9.1.3 (SAS Inc., Cary, NC) and all found to be normally distributed. The means, range of means and the standard errors of Brix, pol and sucrose across the years 2004 and 2005 were calculated from the complete unadjusted data. The analysis of variance (ANOVA) was performed using PROC GLM of SAS using the model $y_{ij}= \mu+\alpha_i+\beta_j+e_{ij}$, where y_{ij} is the phenotypic mean of a particular trait with μ as the overall mean, α_i as the genotype effect, β_j as the block (replication) effect and e_{ij} as the error effect. All effects in the model were considered as random. Estimates of variance components σ^2_g (genotypic variance) and σ^2_e (error variance) were calculated as described by Searle (1971). Broad sense heritability estimates were calculated using the formula $H^2 = \sigma^2_g/(\sigma^2_g + \sigma^2_e/r)$ where σ^2_e is the error variance and r is the number of replications.

Phenotypic (r_p) and genotypic (r_g) correlation coefficients among the traits were calculated based on adjusted means using the PLABSTAT software (Utz, 2001).

6.2.5 QTL Analysis

The method of composite interval mapping (CIM) (Zeng, 1994; Jansen and Stam, 1994) was employed to detect QTLs and to estimate their effects. Cofactors were selected using a stepwise regression method and the final model was selected where the AIC (Akaike's information criterion) was minimized at penalty = 3.0. The phenotypic variation explained by each QTL was determined as the square of the partial correlation coefficient (partial R^2), keeping the effects of all other QTLs for that particular trait as fixed. The proportion of total phenotypic variation explained by all QTLs in the model was determined (adjusted R^2) after making adjustments for number of terms in the multiple regression model. The threshold LOD score for each trait was computed as described by (Churchill and Doerge, 1994) using 1000 permutations. All the QTL analyses were performed using the PLABQTL software (Utz and Melchinger, 1996).

To detect epistatic interactions, every single significant QTL detected from CIM-QTL mapping was tested for linear x linear digenic interactions by multiple regression analysis using SAS ver 9.1.3 (SAS Inc., Cary, NC). The phenotypic variation explained by individual interacting QTLs was determined as partial R^2 and total phenotypic variation accounted by all interacting QTLs was determined as adjusted R^2 of multiple regression. All digenic interactions were detected at $P < 0.05$ level.

6.2.6 Discriminant Analysis

Discriminant analysis was performed according to Mcharo et al. (2004). The population was divided into three groups (high, intermediate and low) based on the Brix, sucrose and pol data using 2 standard deviation grouping. Using the PROC STEPDISC

option of SAS, a forward method parametric discriminant analysis was performed with criteria set to default (SLENTRY = 0.15) to select the most informative markers that differentiates the individuals in the groups. Using PROC DISCRIM, a non-parametric discriminant analysis was performed employing the selected markers to construct and validate a class prediction function and to predict group membership. The cross validation procedure tests for the efficiency of the selected marker derived classification model. The classification error rates derived from the cross validation procedure provide a measure of overall model efficiency. While performing the DA, we assumed that there is no population structure in the mapping population as all the progeny were full sibs.

6.3 Results

6.3.1 Phenotypic Evaluation

The analysis of variance results indicated that there was a significant difference ($P < 0.01$) among the genotypes in the F_1 population for the traits Brix, sucrose and Pol in the years 2004 and 2005 (Fig 6. 1; Table 6. 1). In the year 2004 early Brix estimates ranged from 3.75 to 11.73 with a mean of 7.90. In the same year, late Brix estimates ranged from 6.52 to 11.46 with a mean of 9.24, late sucrose estimates ranged from 2.57 to 9.92 with a mean of 5.89 and late Pol estimates ranged from 14.31 to 47.23 with a mean of 31.73. In 2005, the early Brix estimates ranged from 5.56 to 8.98 with a mean of 6.85, early sucrose estimates ranged from 1.02 to 5.00 with a mean of 2.17 and early Pol estimates ranged from 6.50 to 20.62 with a mean of 11.12.

The broad sense heritability estimates (H^2) in 2004, for early Brix was 0.69 whereas that of late Brix, late sucrose and late Pol were 0.86, 0.85 and 0.90, respectively. In 2005, the H^2 estimates for early Brix, early sucrose and early Pol were 0.75, 0.71 and 0.70, respectively (Table 6. 1). The genotypic coefficient of variation ranged from 0.39

Table 6. 1. The means± SE, ranges of means, ANOVA and broad sense heritability estimates for Brix (B), sucrose (S) and pol (P) traits at early (E) and late (L) seasons in the years 2004 (04) and 2005 (05).

Trait	Mean ± SE	Range	MS	F value	GCV†	H^2
04EB	7.90±0.07	3.75-11.73	2.55	2.76**	8.88	0.69
04LB	9.24±0.05	6.52-11.46	1.57	6.82**	5.64	0.87
04LS	5.89±0.06	2.57-9.92	1.99	7.56**	11.61	0.88
04LP	31.73±0.58	14.32-47.23	79.79	10.15**	0.93	0.90
05EB	6.85±0.03	5.56-8.98	0.53	4.11**	2.22	0.77
05ES	2.17±0.03	1.02-5.00	0.60	3.50**	6.97	0.72
05EP	11.12±0.25	6.50-20.62	17.63	3.93**	0.39	0.74

** Significantly different at 0.01; † Genotypic coefficient of Variation

(05EP) to 11.61 (04LS) and comparable to those observed by Milligan et al. (1992).

Positive genotypic correlation coefficients were observed among all the traits under study (Table 6. 2).

Table 6. 2. Genotypic Correlation coefficient estimates for Brix (B), sucrose (S) and pol (P) estimated at early (E) and late (L) crop growth periods across the years 2004 (04) and 2005 (05).

	04EB	04LB	04LS	04LP	05EB	05ES
04LB	0.74**					
04LS	0.79**	0.96**				
04LP	0.79**	0.90**	0.97**			
05EB	0.93**	0.64**	0.62**	0.62**		
05ES	0.90**	0.76**	0.83**	0.82**	0.81**	
05EP	0.83**	0.72**	0.81**	0.79**	0.98**	0.98**

**Significant at 0.01 level

6.3.2 QTL Analysis

6.3.2.1 *Saccharum officinarum*

The list of QTLs detected in *S. officinarum* is presented in Table 6. 3. A total of fifty QTLs were observed combined across all traits with the LOD scores ranging from

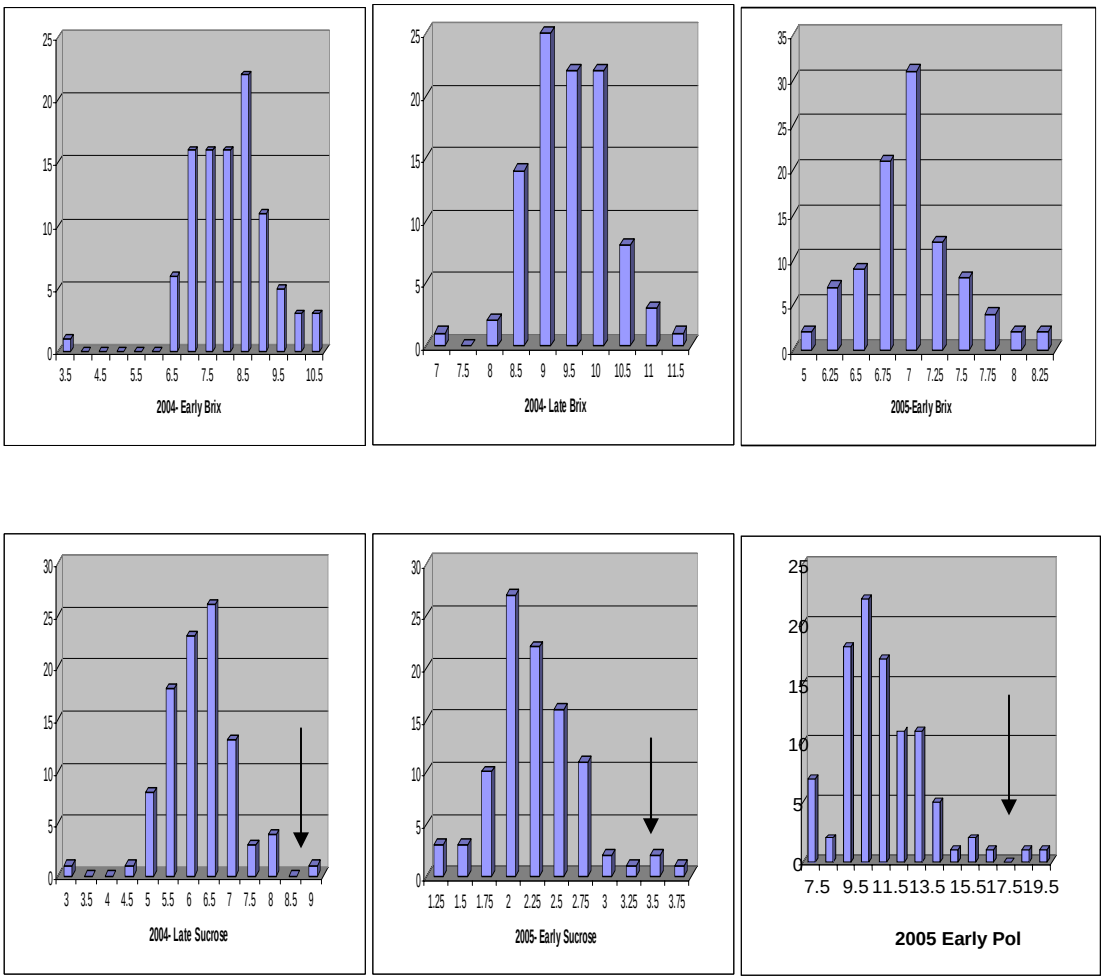


Figure 6. 1. The normal distribution of Brix, pol and sucrose data among the 100 progeny measured at early and late seasons of crop growth during the years 2004 and 2005. The arrow indicates the measurements of Brix or sucrose for *S. officinarum* 'Louisiana Striped'. The *S. spontaneum* data was unavailable as USDA-APHIS regulations do not permit the planting of weedy species in field trials.

2.51 to 7.64. In the year 2004, eight QTLs were observed for 04EB on the linkage groups (LG) L6, L8, L15, L35, L40 and L49 explaining a total variation of 48.4%. The variation explained by the individual QTL ranged from 1.2% (AF11715 on LG L15) to

34.3% (AF1186 on LG L6). For 04LB, eight QTLs were observed which explained a total variation of 35.5%. The variation explained by individual 04LB QTLs ranged from 0.2% (AF11715 on LG L15) to 27.2% (AF1186 on LG L6). Two QTLs on LG L6 and one each on LG L8 and L49 were commonly observed for both 04EB and 04LB. For the trait 04LS, fourteen QTLs were detected accounting for a total variation of 28.8%. The variation accounted for by individual 04LS QTLs ranged from 0.1% (AF886 on LG L20) to 22.7% (AF475 on LG L8). Four QTLs were observed for 04LP on LGs L3, L4, L6 and L36 accounting for a total variation of 15.9%. The variation explained by individual QTL ranged from 7.3% (AF7710 on L3) to 12.5% (AF274 on L36). The QTL on LG L20 and L49 was common for 04LB and 04LS whereas the QTLs on LG L8 and L15 were common for 04EB, 04LB and 04LS traits. The negative effect QTL on LG L6 (AF1186) was consistently present in all the traits measured in 2004.

In 2005, eight QTLs (two on LG L6 and L15 and one each on LG L8, L35, L40 and L49) were observed for 05EB explaining a variation of 47.4%. The variation explained by individual 05EB QTLs ranged from 0.3% (AF11715 on LG L15) to 32.2%

Table 6. 3. The markers identified in *S. officinarum* 'Louisiana Striped' based on composite interval mapping (CIM) QTL analysis in an interspecific *S. officinarum x S. spontaneum* cross.

Trait	LG[†]	Marker	LOD score[‡]	partial R^{2}[§]	Additive effect
04EB	L6	AF1286	4.62	21.2	0.109
		AF1186	7.48	34.3	-0.168
	L8	AF475	4.42	11.3	0.058
	L15	AF11714	4.56	3.0	0.053
		AF11715	2.92	1.2	0.013
	L35	AF287	2.92	15.1	-0.068
	L40	AF486	4.60	17.4	0.092
	L49	PM05810	4.56	15.9	0.088
				Adj.R^{2}[¶] = 48.4	
04LB	L1	AF5811	2.77	0.5	0.114
	L3	AF7717	4.32	9.8	-0.238
	L6	AF1286	5.07	6.9	0.258
		AF1186	7.05	27.2	-0.397
	L8	AF475	3.11	12.1	-0.135

Cont'd.

	L15	AF11716	3.25	0.2	0.119
	L20	AF886	3.44	4.4	0.150
	L49	PM05810	2.76	15.1	0.152
				Adj.R^2 = 35.5	
04LS	L1	AF5814	4.84	14.3	-0.364
		AF588	5.25	9.6	0.390
		AF5717	2.68	8.0	0.161
	L5	AF4811	3.49	5.2	-0.196
	L6	AF1286	3.50	11.4	0.167
		AF1186	7.36	21.9	-0.318
	L8	AF475	4.14	22.7	-0.191
	L15	AF11716	2.56	0.2	0.096
	L20	AF886	3.32	0.2	0.081
	L25	AF1377	2.97	10.2	-0.123
	L30	sr329	3.14	0.2	0.144
	L33	sus41	2.61	2.8	-0.177
	L37	PM0882	3.80	10.3	-0.214
	L41	PM06714	2.51	4.4	0.150
				Adj.R^2 = 28.4	
04LP	L3	AF7710	2.73	7.3	-1.8
	L4	AF12811	3.44	7.4	2.00
	L6	AF1186	3.21	9.8	-2.08
	L36	AF274	2.58	12.5	2.05
				Adj.R^2 = 15.9	
05EB	L6	AF1286	4.08	21.6	0.114
		AF1186	6.35	32.2	-0.173
	L8	AF475	4.22	11.0	-0.061
	L15	AF11714	4.37	1.5	0.061
		AF11715	3.66	0.3	0.022
	L35	AF287	2.65	16.7	-0.049
	L40	AF486	3.88	19.8	0.095
	L49	PM05810	2.98	14.4	0.077
				Adj.R^2 = 47.4	
05ES	L6	AF1286	5.59	23.5	0.127
		AF1186	7.64	30.9	-0.175
	L8	AF475	3.22	12.0	-0.055
	L15	AF11714	4.87	4.4	0.057
		AF11717	3.03	4.3	0.026
	L35	AF287	3.24	17.5	-0.059
	L40	AF486	3.86	17.0	0.090
	L49	PM05810	4.57	15.9	0.092
				Adj.R^2= 47.8	
				Avg. Adj.R^2 = 41.4	

[†]The construction of linkage groups (LG) was described in Chapter 5.
[‡]The threshold LOD score was 3.01 as detected by 1000 run permutation test and the QTLs with > 3.01 LOD score were deemed putative.
[§]Proportion of phenotypic variation explained by individual QTL.
[¶]Proportion of total phenotypic variation explained by all QTL in the final model after adjusting for number of terms in multiple regression.

(AF1186 on LG L6). For 05ES, the same QTLs were observed except for a different marker on LG L15 (AF11717) explaining a total variation of 47.8%. The variation accounted by individual QTLs ranged from 4.3% (AF11717 on LG L15) to 30.9% (AF1186 on LG L6). No QTL were detected for 05EP.

Combined across both years, the QTL on LG L6 and L8 were consistently detected for all the traits whereas the QTL on L49 was observed for all but the trait 04LS. The QTL on LG L40 was consistently observed for 04EB, 05EB and 05LB. Most of the QTLs had positive additive effect except for a few which had negative effects.

6.3.2.2 *Saccharum spontaneum*

The list of QTLs detected in *S. spontaneum* is presented in Table 6. 4. A total of 26 QTL were observed with the LOD scores ranging from 2.56 to 7.59. In 2004, three QTLs were observed for 04EB on LG S5, S6 and S33 explaining a total phenotypic variation of 31.3%. The variation explained by individual QTLs ranged from 0.2% (AF2711 on LG S6) to 7.0% (PM0972 on LG S5). Five QTLs (on LG S5, S6, S12, S21 and S36) were observed for 04LB explaining a total variation of 43.0%. The variation accounted by individual QTLs ranged from 0.1% (sr424 on LG S21) to 15.1% (PM0983 on LG S5). For the trait 04LP, three QTL were observed explaining a total variation of 6.5%. The variation explained by individual QTL ranged from 1.8% (cd55 on LG S4) to 7.8% (AF1477 on S33). For the trait 04LS, five QTLs (three on LG S8 and one each on LG S3 and S35) were observed explaining a total variation of 23.6%. The variation accounted for by individual 04LS QTLs ranged from 2.9% (AF577 on LG S8) to 17.9% (AF575 on LG S8). In 2004, the QTL (AF2711) on LG S6 was common for 04EB and 04LB whereas the QTL on S33 (AF1477) was common for 04EB and 04LP.

Table 6. 4. The markers identified in *S. spontaneum* 'SES 147B' based on composite interval mapping (CIM) QTL analysis in an interspecific *S. officinarum x S. spontaneum* cross.

Trait	LG[†]	Marker	LOD score[‡]	partial $R^{2§}$	Additive effect
04EB	S5	PM0972	3.56	7.0	0.122
	S6	AF2711	3.57	0.2	-0.023
	S33	AF1477	3.77	4.8	-0.093
				Adj.$R^{2¶}$=31.3	
04LB	S5	PM0983	3.97	15.1	0.308
	S6	AF2711	3.30	0.3	-0.005
	S12	sr416	2.62	8.1	-0.140
	S21	sr424	2.88	0.1	0.048
	S36	PM0781	2.96	3.8	0.115
				Adj.R^2=43.0	
04LS	S3	AF888	3.31	10.7	0.247
	S8	AF575	5.87	17.9	-0.465
		AF577	2.86	2.9	-0.187
		AF589	3.44	15.4	0.355
	S35	PM0886	2.56	9.0	0.225
				Adj.R^2=23.6	
04LP	S4	cd55	2.87	1.8	1.8
	S21	sr424	3.07	6.4	-2.0
	S33	AF1477	2.61	7.6	-1.4
				Adj.R^2=6.5	
05EB	S6	AF2711	3.31	0.1	-0.010
	S33	AF1477	2.91	2.7	-0.073
				Adj.R^2=23.0	
05ES	S5	PM0981	3.65	6.2	-0.125
		PM0983	7.59	25.6	0.254
	S6	AF2711	3.47	0.6	-0.009
	S8	AF575	3.10	9.0	-0.129
	S30	AF979	2.61	5.6	-0.103
	S33	AF1477	3.62	10.5	-0.097
	S38	PM0375	3.25	11.0	0.145
				Adj.R^2=43.5	
05EP	S1	AF11720	3.19	9.0	-1.02
				Adj.R^2=7.0	
				Avg. Adj.R^2 = 41.4	

[†]The construction of linkage groups (LG) was described in Chapter 5.
[‡]The threshold LOD score was 2.90 as detected by 1000 run permutation test and the QTLs with > 2.90 LOD score were deemed putative.
[§]Proportion of phenotypic variation explained by individual QTL.
[¶]Proportion of total phenotypic variation explained by all QTLs in the final model after adjusting for number of terms in multiple regression.

In 2005, two QTLs were observed for 05EB (on LG S6 and S33) whereas seven QTLs were observed for 05ES (two on LG S5 and one each on LG S6, S8, S30, S33 and

S38). The phenotypic variation explained by 05EB QTLs and 05ES QTLs was found to be 23.0% and 43.5%, respectively. The variation explained by individual 05EB QTLs ranged from 0.1% (AF2711 on LG S6) to 2.7% (AF1477 on LG S33) whereas that of 05ES QTLs ranged from 0.6% (AF2711 on LG S6) to 11% (PM0375 on LG S38). Only one QTL was observed for 05EP on LG S1 (AF11720) explaining a variation of 7.0%. The QTLs on LG S6 and S33 were commonly observed for all the traits except for 04LS. Most of the QTLs had negative effects except for a few QTL which had positive additive effects.

Table 6. 5. The linear x linear digenic interacting QTL (*i*QTL) detected in *S. officinarum* 'Louisiana Striped' and *S. spontaneum* 'SES 147B' parental species.

Trait	Number of interacting QTL[†]	Range of partial $R^{2‡}$	Adjusted $R^{2§}$	Range *P*[¶] values
		S. officinarum		
04EB	1	2.4	2.4	0.01
04LB	6	2.1 – 12.5	28.6	0.03 – 0.0008
04LS	14	1.8 – 6.9	53.3	0.03 – 0.0001
04LP	1	6.8	6.8	0.01
05EB	8	1.4 – 8.0	38.5	0.02 – 0.0001
05ES	4	2.5 - 8.0	23.2	0.03 – 0.0001
Total	34			
		S. spontaneum		
04EB	1	4.6	4.6	0.05
04LB	1	4.7	4.7	0.05
04LP	1	4.8	4.8	0.04
05ES	1	4.9	4.9	0.03
Total	4			

[†]Linear x Linear digenic interacting QTLs as observed from multiple regression analysis.
[‡]Proportion of phenotypic variation explained by individual interacting QTL.
[§]Proportion of total phenotypic variation explained by all interacting QTL in the final model after adjusting for number of terms in multiple regression.
[¶]The *P* values are significant at 0.05 level.

6.3.2.3 Digenic Interactions

The number of linear x linear digenic interactions detected in *S. officinarum* ranged from 1 (in 04EB and 04LP) to 12 (in 04LS) with a total of 34 interacting QTLs

(*i*QTL) (Table 6. 5). The total phenotypic variation accounted by all *i*QTL within a trait ranged from 2.4% (04EB) to 53.3% (04LS). Most of the 34 *i*QTLs were detected only once except for two: *i*QTL AF11715/PM05810 was detected in 04EB and 05EB and *i*QTL AF475/AF11714 was detected in 05EB and 05ES. Four digenic *i*QTL were detected in *S. spontaneum*, one each in 04EB, 04LB, 05ES and 04LP and none were observed in 04ES and 05EB (Table 6. 5). All the *i*QTLs involved either one of these two markers: AF2711 on LG S6 and AF1477 on LG S33. The total phenotypic variation explained by the *i*QTL was minimum in 04EB (4.6%) and maximum in 05ES (4.9%).

6.3.3 Discriminant Analysis

The genotypes in the population were divided into three groups based on Brix, sucrose and/or pol data (low, medium and high) using a 2-standard deviation (2SD) differentiation. Assuming no population structure (as all the genotypes originated from the same cross), a minimum of 10 markers (for each trait) detected by the discriminant analysis procedure gave > 90% classification (< 10% error rate) of the genotypes with both *S. officinarum* and *S. spontaneum* data sets. Over and above 10 markers, a 100% classification was consistently achieved for all the traits using the data sets from both parental species.

Several of the DA-identified markers were found to be common to Brix, sucrose and pol in the early as well as in late plant growing seasons and across both crop-years. In *S. officinarum*, the marker sr546 was common to 04EB and 05ES, AF1182 was common to 04EB, 04LS and 04LP and the marker AF5819 was common to 04LS and 05EP. Two markers sr2211 and AF1572 were consistently detected for 04EB and 05EB (Table 6. 6).

Table 6. 6. The markers identified in *S. officinarum* 'Louisiana Striped' from discriminant analysis based on 2 standard deviation (SD) differentiation

Year-trait	DA selected markers	Percent classification for DA identified markers		
		15	10	5
04EB	***sr2211***, sr572, **sr546,** sr124, AF5810, AF287, **AF1182**, AF1273, AF1572, cd23	100	99.98	82.50
04LB	sr259, sr576, AF473, AF779, AF1186, AF1277, cd91, PM0171, PM0376, PM0383	100	99.98	72.58
04LS	sr622, sr632, sr612, sr132, *AF5819*, **AF1182**, AF1184, cd144, PM08811, PM08713	100	99.97	75.42
04LP	sr329, AF572, AF986, **AF1182**, AF12715, AF12720, AF13718, AF14811, cd143, PM02817	100	99.35	80.06
05EB	***sr2211***, AF774, AF578, AF279, AF12716, AF1487, AF1572, cd59, cd141, sr624	100	100	68.53
05ES	sr229,**sr546**, sr634, sr643, AF5713, AF9713, AF1485, sus41, PM02712, PM03712	99.99	99.98	68.83
05EP	PM0374, AF1476, AF1575, AF1573, AF1288, AF11718, AF988, AF7714, *AF5819*, AF588	100	100	84.91

†The markers which were repeatedly detected for any two traits are represented either in bold or italics or underlined or combinations of these.

Likewise, in *S. spontaneum,* three markers sr157 and cd54 were common to 04EB and 05EB whereas the marker AF872 was common to in 04EB, 04LP and 05EB. Two markers sr381 and PM0586 were detected for 04LB and 04LS whereas the marker PM0284 was detected for 04EB, 04LS and 05ES. The marker AF1272 was common to 04LS and 05EB, cd54 was common to 04EB and 05EB, cd56 was common to 05ES and 05EP and the marker sus43 was common to 04LP and 05EB (Table 6. 7).

6.4 Discussion

6.4.1 Marker Detection

In a traditional quantitative trait loci (QTL) analysis, the ability to detect QTLs depends in part on the magnitude of QTL effects and on the structure of the mapping

Table 6. 7. The markers identified in *S. spontaneum* 'SES 147 B' from discriminant analysis based on 2 standard deviation (SD) differentiation.

Year-trait	DA selected markers†	Percent classification for DA identified markers		
		15	10	5
04EB	sr356, ***sr157***, ***AF872***, AF873, AF273, **cd54**, sai34, **PM0284**, PM0384, PM06711	100	99.98	82.50
04LB	sr344, sr381, AF975, AF1172, AF1173, sai24, PM0883, PM0678, PM0586, PM0372	100	99.98	72.58
04LS	sr333, sr341, *sr532*, sr556, sr641, sr381, *AF1272*, cd27, PM0586, **PM0284**	100	99.97	75.42
04LP	***AF872***, AF2712, AF14710, **sus43**, PM0176, PM0686, PM0586, cd53, PM03711, PM0577	100	100	94.31
05EB	sr325, ***sr157***, sr126, ***AF872***, AF11712, *AF1272*, AF1473, **sus43**, **cd54**, PM02714	100	100	68.53
05ES	sr4210, sr566, AF771, AF282, cd21, sus21, **cd56**, PM0177, PM0684, **PM0284**	99.99	99.98	68.83
05EP	sr442, sr646, sr661, *sr532*, AF2711, AF1174, **cd56,** PM0972, PM0887, PM0276	100	100	83.65

†The markers which were repeatedly detected for any two traits are represented either in bold or italics or underlined or combinations of these.

population. In sugarcane, using relatively larger mapping populations (> 200) and number of linked markers, Hoarau et al. (2002) and Aitken et al. (2006) detected several QTLs for agronomic traits including sugar content with individual QTL effects ranging from 3 to 7% and 3 to 9%, respectively. Hoarau et al. (2002) used a selfed population of a cultivar R570 and Aitken et al. (2006) used a cultivar x *S. officinarum* cross. Although the population sizes in these studies were comparatively large, the mapping populations were derived from closely related, high sugar producing genotypes. Due to high selection pressure over several decades of recurrent breeding, favorable alleles get accumulated in the cultivars which potentially minimize genetic variation and the ability

to distinguish trait segregation. Therefore, most of the QTLs detected in these studies had minor effects compared to our study.

Our study used a mapping population derived from a interspecific (*S. officinarum* x *S. spontaneum)* cross. This population, although atypical of populations used in breeding, was created following the objectives of the Louisiana sugarcane improvement program which include identification of favorable genes from the basic *Saccharum* germplasm. The phenotypic variation explained by individual QTLs (partial R^2) was found to be as high as 23.5% (AF1286 on L6 for 05ES) in *S. officinarum* and 25.4% (PM0983 on S5 for 05ES) in *S. spontaneum* with positive additive effects. In *S. officinarum*, most (65%) of the QTLs detected had a positive effect on sugar accumulation although a few (35%) QTLs having negative effects were detected. Conversely, most (65%) of the QTLs contributed by the *S. spontaneum* parent had a negative effect on sugar accumulation but remarkably, some QTLs having a positive effect were detected. In both species the QTL effects remained consistent for all the traits regardless of season (early vs. late) and crop-year (04 vs. 05). Previous sugarcane studies have reported consistency of QTL effects for Brix and pol across seasons (Aitken et al., 2006) and crop-years (Aitken et al., 2006; Hoarau et al., 2002).

With a population size of 100 individuals, we could detect QTLs with large effects, but perhaps minor QTLs went undetected due to the relatively small population size and low marker density in the linkage map. However, even with a relatively smaller population size of 44 F_1 individuals derived from a *S. officinarum* x *S. robustum* cross, Sills et al. (1995) detected several QTLs with relatively large (23-58 %) effects. However, unlike in the Sills et al. (1995) study where single factor ANOVA approach was used, we used the more reliable approach of composite interval mapping (CIM) QTL

analysis (Zeng et al., 1994). The results from CIM-QTL method are usually comparable to those obtained from Mapmaker/QTL analysis (Al-Janabi et al., 2007). Nevertheless, a much larger population size and dense map is necessary to better appreciate the number of loci governing quantitative traits.

The sucrose and the traits from which it is estimated, Brix and pol, are all moderately to highly heritable traits (Kang et al., 1983; Milligan et al., 1990 and 1992; Hoarau et al., 2002; Aitken et al., 2006). As such, one would expect these traits to be governed by one or a few major QTL or genes. However, these heritability estimates are based on the total genetic variance (broad sense heritability) whereas markers account for only additive but not non-additive effects (Sills et al., 1995). Therefore, it is not surprising that in this and several previous studies (Ming et al., 2001; Hoarau et al., 2002; Aitken et al., 2006) minor QTLs were found to be associated with these traits. Thus, although we found some QTLs with large effects the importance of QTLs with small effects cannot be undermined (Ming et al., 2001).

Measurements of Brix and pol are needed to estimate sucrose content. Brix is easier to measure (g solute per 100 g of solution) whereas pol (g sucrose/ 100g of juice) is harder, more expensive, less environmental friendly and time consuming. As an indirect measure of sucrose content, Brix is especially important during the early stages of breeding program. During this stage, breeders evaluate a prohibitively large number of clones and time and the cost of measuring for sucrose can serve as limiting factors. In this study significant ($P < 0.01$) positive correlations were found between Brix, pol and their derived trait sucrose content (Table 6. 2). Although several common QTLs were expected among these traits, surprisingly, few (AF1186 on LG L6; sr424 on LG S21 and AF1477 on LG S33) were found. Also QTLs were found for sucrose content that could

neither be attributed to Brix nor pol. This disparity might partially be due to the fact that fewer QTLs were found for pol compared to Brix. The QTLs for Brix possibly included other solutes including sucrose. Moreover the correlations were based on total genetic variance whereas the QTLs measured only the additive portion of the variation.

Early harvesting of sugarcane would have remarkable benefits for the sugar industry. In Louisiana, initiating crushing earlier in the season would help the crop escape cold temperatures that occur later in the season (November- December). In order to initiate early harvesting, however, the genotype(s) should accumulate sucrose at levels high enough for commercial exploitation. For a successful and profitable implementation of early harvesting, therefore, reliable information is required on absolute levels of sucrose at the early and at normal harvesting periods along with cane yield and ratoon performance (Cox, 1999). Cox et al. (1990) asserted that there might be certain genes which are up-regulated only during early plant growing season and *vice versa*. In this study, two positive effect markers/QTLs were unique to early season Brix/sucrose (AF486 on LG L40 and PM05810 on LG L49). Such markers might potentially aid in selecting for high early sugar accumulating clones.

Jackson and Morgan (2003) showed that it is possible to select early for the genotypes which prove to be competent with enough sucrose levels at the late season provided there is no genotype-season interaction. They advocate selecting early for sucrose accumulation so that 1) attention and resources can be focused on other traits later in the season or 2) the clones can be replanted within the same growing season to save a year of selection. In this study, significant positive correlations were observed for early and late season within a crop-year (r_g= 0.74 between 04EB and 04LB; r_g= 0.79 between 04EB and 04LS) and across crop-years (r_g = 0.49 between 05EB and 04EB; r_g =

0.55 between 05EB and 04ES; r_g = 0.64 between 04LS and 05EB; r_g = 0.82 between 04LS and 05ES; r_g = 0.64 between 04LB and 05EB and r_g = 0.76 between 04LB and 05ES). The positive correlations were supported further by the presence of common QTLs with positive effects for early as well as late season (AF1286 on LG L6 and PM05810 on LG L49 in *S. officinarum*) sucrose content. It was also observed that there was no rank change in most of the clones for sugar content increase from early to late season i.e., the high early sugar genotypes tended to be high late sugar genotypes. The positive correlations and common markers for both early and late season observed in this study corroborates the assertions from other studies (Jackson and Morgan, 2003; Aitken et al., 2006) stating that it is possible to select promising genotypes early in the season.

6.4.2 Consistency of the Detected Markers

The consistency of marker-QTL associations in different populations across different locations and years is the key to successful masker assisted selection (MAS) breeding. In this study, four QTLs in *S. officinarum* (AF1286 on LG L6, AF11714 on LG L15, AF486 on LG L40 and PM05810 on LG L49 (AF475 on L8) and one QTL in *S. spontaneum* (PM0983 on LG S5) were consistently observed in 2004 and 2005. In addition, high significant correlations coupled with high heritabilities in both crop-years and between early and late season sugar accumulation indicate that common genes might be governing these traits. However, the low numbers of common QTLs observed across crop-years could be due to genotype-year interactions as noticed for most quantitative traits in sugarcane (Kang et al., 1987; Jackson and Hogarth, 1992; Gravois et al., 2002). Nevertheless, the effects of these common markers across crop-years were remarkably in the same direction. The common crop-year markers with positive effects could be regarded as strong marker-QTL associations and could potentially be useful for MAS. In

a clonally propagated crop like sugarcane once a strong marker-QTL association is detected in a progeny population, it has an immediate role in crop improvement via clonal selection as there is no further probability of cross-over between the marker and the QTL.

6.4.3 Digenic Interactions

The digenic epistatic interactions (*i*QTL) explained significant amount of phenotypic variation for all of the traits in both *Saccharum* parental species. However, the number of detected *i*QTLs were not consistent in both years as previously observed (Hoarau et al. 2002; Aitken et al. 2006). Since the epistatic interactions were tested only among the detected QTLs, one might assume that several other interactions between non-QTLs went undetected. Studies in other crops have detected interactions involving QTLs and non-QTLs (Kulwal et al., 2004; Li et al., 1998). The notable aspect of most of the digenic interactions observed in both parental species is that they involved either *Pst* I-*Mse* I or SRAP (in *S. spontaneum*) and/or TRAP (*S. officinarum*) markers consistent with the observation that these markers are amplified from actively transcribing genic regions (which is discussed in the following part). In *S. officinarum*, the detected *i*QTL mostly had *Pst* I – *Mse* I derived AFLP markers. Interestingly, out of 14 *i*QTL found in 04LS, 7 *i*QTL involved a TRAP derived marker (sus41). As an independent main effect marker-QTL, although sus41 surprisingly had negative effect on sucrose content, the *i*QTL involving sus41 were mostly positive. In *S. spontaneum*, out of three digenic interactions detected one involved a *Pst* I – *Mse* I marker and other involved a SRAP marker. All the *i*QTL effects in *S. spontaneum* were found to be positive.

6.4.4 Comparison among Markers

*Eco*R I – *Mse* I derived AFLP markers have been widely used in many crops for linkage mapping and QTL analysis including sugarcane (Hoarau et al., 2001; Aitken et al., 2005). However, most of the *Eco*R I – *Mse* I derived polymorphisms are randomly distributed across the genome and methylation insensitive *EcoR*-I enzyme cannot differentiate between genic and non-genic regions. It was reported that the actively transcribing genic regions in the genome are generally hypomethylated (Barret and Kidwell, 1998) and therefore in this study, we have included a methylation sensitive *Pst* I enzyme along with an *EcoR*-I enzyme. The two *Pst* I- *Mse* I derived QTLs in *S. officinarum* (PM06714 on LG L41 and PM05810 on LG L49) and four in *S. spontaneum* (PM0983 on LG S5, PM0886 on LG S35, PM0781 on LG S36 and PM0375 on LG S38) represent hypomethylated regions. However, the assertion that these markers represent genes in the sucrose pathway genes can only be verified by sequencing the bands and BLASTing for homologous sequences.

Several studies have previously documented that SRAP and TRAP techniques target genic regions of the genome to detect polymorphisms (Li and Quiros, 2001; Liu et al., 2005; Alwala et al., 2006). In this study, a few QTLs associated with SRAP (sr329 on LG L30 in *S. officinarum*; sr424 on LG S21 in *S. spontaneum*) and TRAP markers (sus41 on LG L33 in *S. officinarum*) were detected. Since, most of the single dose SRAP and TRAP markers were not linked, most of the other associated QTLs might have gone undetected. In a crop like sugarcane with a large genome size, a large number of markers (such as those generated by AFLP) are required and SRAP and TRAP markers could be used to complement the already existing linkage maps since the accuracy of QTL detection is directly proportional to the map saturation. Nevertheless, the basic

advantages of using SRAP and TRAP technique is that they target genic regions of the genome (Li and Quiros, 2001; Alwala et al., 2006) and therefore, they pose immediate applications in MAS. Although SRAP markers have been suggested to play a major role in marker assisted selection, their utility in detecting QTL has not been reported in many crops including sugarcane. On the other hand, TRAP markers have been effectively employed to tag genes for important agronomic traits in wheat (Liu et al., 2005), disease resistance traits in common bean (Miklas et al., 2006) and insect resistance trait in wheat (Wang et al., 2006). This is the first report identifying the SRAP and TRAP marker based QTL and *i*QTL associated with sucrose genes in sugarcane.

6.4.5 Implications in Introgression Breeding

One of the potential reasons for the limited progress in sugarcane improvement is the genetic 'bottleneck' in the development of cultivars. The current trend of making crosses between high sucrose producing parental cultivars/clones followed by selection of high sugar producing progeny has lead to further narrowing of the genetic base in sugarcane. A study from our lab to assess the genetic variability and population structure among different clones/cultivars has documented that the genetic base of US sugarcane germplasm is narrow (Arro, 2005) corroborating a previous study (Deren, 1995). The narrow genetic base coupled with high selection pressure could have lead to accumulation of most of the favorable alleles in current sugarcane cultivars, thus limiting the expansion of genetic variation. Up to 80 % of the genome of sugarcane cultivars is made up of the high sugar producing species *S. officinarum* with 15 % represented by the wild *S. spontaneum* and 5 % by recombinant chromosomes (D'Hont et al., 1996). Yet, most of the diversity found among sugarcane cutivars could be attributed to *S.*

spontaneum (Jannoo et al., 1999; Reffay et al. 2005). Therefore, *S. spontaneum* is being considered as a potential untapped source for novel alleles for sugarcane improvement.

Tanksley and McCouch (1997) affirmed that wild germplasms harbor a huge cache of agronomically important genes which could be exploited and introgressed into the cultivated background to further enhance the productivity of crop plants. In their study, in tomato, when QTLs from wild species *Lycopersicon hirsutum* were transferred, the lines outperformed the elite varieties (Bernacchi et al., 2004). Similarly, when genes from wild ancestor *L. pimpinellifolium* were introduced, the fruit size in the cultivated tomato was dramatically increased (Tanksley et al., 1996). In sugarcane, therefore, it is highly probable that the presence of only 15% of *S. spontaneum* genome might have contributed some of the favorable sucrose alleles in the current cultivars. By retracing the movement of alleles from *S. spontaneum* through the breeding program, Reffay et al. (2005) successfully showed that alleles from this wild relative had contributed favorably to sucrose content in cultivated sugarcane. In our study, in addition to detecting several positive QTL from *S. officinarum*, a few positive QTLs were also detected from *S. spontaneum*. Remarkably, the QTLs from *S. spontaneum* were repeatedly detected across seasons and across crop-years. The positive marker-QTL associations (AF888 on S3, PM0983 on S5, AF589 on S8, sr424 on S21 and PM0375 on S38) from *S. spontaneum* observed in this study could serve as a starting point for MAS in introgression breeding in the Louisiana sugarcane improvement program. One of the *S. spontaneum* marker-QTL (PM0983 on LG S5) which was consistently observed in early and late season and in both crop-years has potential implications in introgression breeding. Besides selecting the positive alleles, marker assisted selection also aids in the purging of negative alleles either from *S. spontaneum* or even from *S. officinarum* (AF1186 on LG L6) mainly to

facilitate accumulation of most of the favorable alleles and maximize the chances of developing improved sugarcane cultivars. Molecular breeding *via* marker assisted selection based introgression breeding provides a better approach to select parental clones or elite clones with enhanced breeding values so as to break the plateau existing for sucrose content in current sugarcane breeding programs.

6.4.6. Discriminant Analysis

DA is an attractive multivariate statistical tool for plant breeders which can differentiate groups of individuals by screening for differences in the variables (or markers), given their quantitative measurements. DA simply finds the linear relationships of quantitative measurements with those of the variables or markers by keeping a minimal variation within the groups of individuals. Although based on several statistical assumptions such as normality of data and homogeneity of covariance matrices DA has proven to be robust with minor violations of these assumptions (Klecka, 1980; Zhang et al., 2005) especially when the marker profile used is categorical data. Recently, using the DA approach, several microsatellite markers associated with agronomic traits were identified in rice (Capdevielle et al., 2002; Aluko, 2003; Zhang et al., 2005) and Alwala et al. (2007) identified AFLP and TRAP markers associated with resistance to *A. flavus* in maize.

In the absence of saturated maps, as in the case of sugarcane, DA provides an attractive platform and a good complementation to QTL analysis to identify markers associated with traits of agronomic interest. Since, we have framework genetic linkage maps for the *S. officinarum* and *S. spontaneum* parents, the DA-identified markers were cross validated with those from QTL analysis. A few of the DA-identified markers were also detected in the QTL analysis (AF1186 on LG L6 and AF287 on LG L35 in

Table 6. 8. The markers identified in discriminant analysis but not in QTL analysis from *S. officinarum* 'Louisiana Striped' and *S. spontaneum* 'SES 147B'.

Year-trait	*S. officinarum*		*S. spontaneum*	
	A[†]	B[‡]	A[†]	B[‡]
04EB	sr546 (L24), AF5810 (L1)	AF1273, sr2211, sr572, sr124, AF1572, cd23	AF872 (S29)	sr356, sr157, AF273, cd54, sai34, PM0284, PM0384, PM06711
04LB	AF473 (L9), PM0376 (L38), PM0383 (L43)	sr259, sr576, AF1277, cd91, PM0171	sr344 (S24), AF1172 (S1), AF1173 (S1)	sr381, AF975, sai24, PM0586, PM0883, PM0678, PM0372
04LS		sr622, sr612, sr632, sr132, AF5819, cd144, PM08811, PM08713	sr333 (S40), sr341 (S44)	sr532, sr556, sr641, sr381, AF1272, cd27, PM0586, PM0284
04LP	sr329 (L30), AF986 (L16), AF12715 (L4), AF13718??,	AF572, AF12720, AF14811, cd143, PM02817	AF872 (S29), PM0686 (S37), PM03711 (S4)	AF2712, sus43, PM0176, PM0586, PM0577, cd53
05EB	AF279 (L12), AF12716 (L4), AF1487 (17), cd59 (L26)	sr2211, AF578, cd141, AF1572, sr624	sr325 (S39), sr126 (S25), AF872 (S29), AF11712 (S1), PM02714 (22)	sr157, AF1272, AF1473, sus43, cd54
05ES	sr546 (L24), sr634 (L44), AF1485 (L18)	sr229, sr643, AF5713, AF9713, PM02712, PM03712	sr4210 (S31), AF282 (S28)	sr566, AF771, cd21, sus21, cd56, PM0177, PM0684, PM0284
05EP	AF1476 (L17)	PM0374, AF1575, AF1573, AF988, AF7714, AF5819	sr442 (S44), sr661 (S45), AF2711, AF1174, PM0972, PM0276	sr646, sr532, cd56, PM0887

[†] The markers identified in DA pointing to new regions on genetic linkage maps which were not identified in QTL analysis.
[‡]The markers identified in DA which were neither detected in QTL analysis nor located to any positions on genetic linkage map.

S. officinarum; Fig 6. 2). Several of the DA-identified markers were localized in the vicinity of QTL-identified markers (in *S. officinarum* marker AF774 is at 6 cM distance from AF7717 on LG L3; marker AF1182 at 20 cM distance from AF1186 on LG L6; marker AF1184 at 15 cM distance from AF11715 on LG L15 and in *S. spontaneum* marker AF873 is at 43 cM from AF888 on LG S3; Fig 6. 2). Additionally, some of the DA-identified markers in *S. officinarum* and in *S. spontaneum* pointed to new genomic regions, which were not identified in QTL analysis (Table 6. 8). Some of the markers which where neither linked in genetic linkage maps of *S. officinarum* and in *S. spontaneum* nor detected by QTL analysis, were also identified by DA representing new loci in the genome affecting the traits. Given the unsaturated linkage maps coupled with high number of unlinked markers (~65% of the markers were unlinked), DA seems to be a fair approach in correctly identifying the markers affecting traits even without *apriori* linkage maps.

Statistical models based on large populations are always expected to be reliable. However, Cruz-Castillo et al. (1994) observed that in DA reliable information might not be extracted if too many markers were identified even with large population sizes. Additionally, they also stated that for DA to be reliable, the population size should be 10 times larger than number of markers to be selected. Using a population size of approximately 50 genotypes per phenotypic group in sweetpotatos, Mcharo et al. (2005) previously identified five to seven markers associated with resistance to root knot nematode. In this study, the population comprised of 100 progeny which were approximately evenly distributed (~33) in each of the three groups (low, medium and high) of genotypes. The highest Wilk's lambda estimate for any marker was found to be 0.79 regardless of trait across both the species and the rest of the estimates were smaller.

Although, QTL analysis identified a few markers with large effects as affecting the trait, given the population size of 100 individuals and 10 markers identified by DA, we are confident that the DA-identified markers are reliable (supported by QTL analysis) and potentially useful in marker assisted selection.

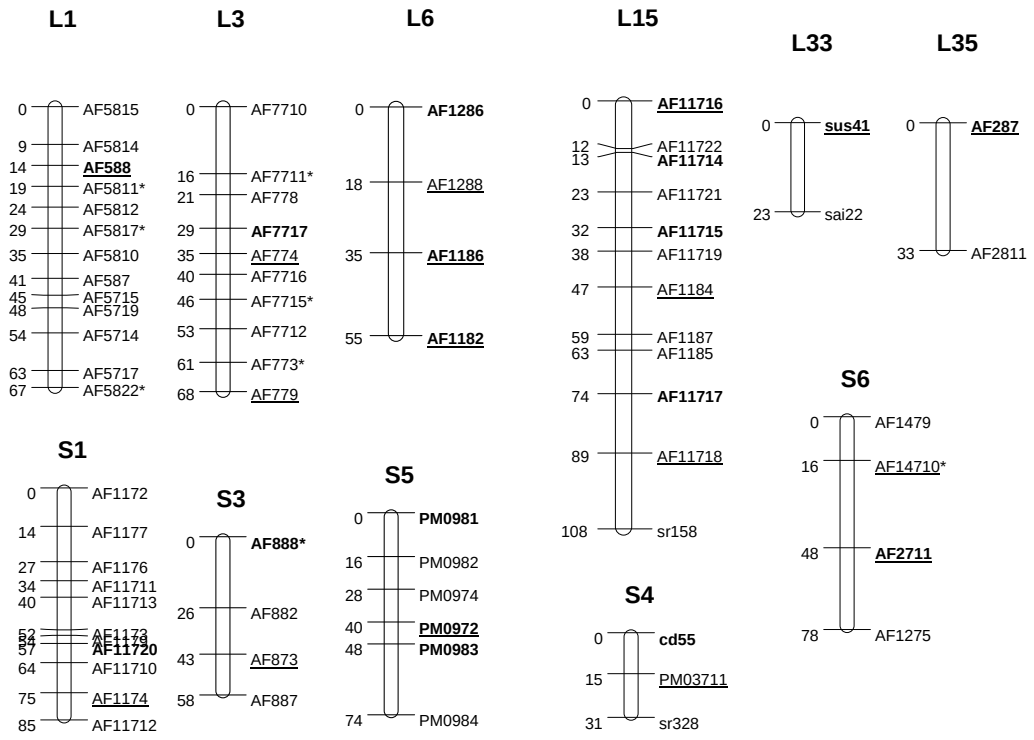


Figure 6. 2. Linkage group locations of markers identified in discriminant analysis (DA) and by composite interval mapping QTL analysis. The DA identified marker are indicated as underlined, QTL identified marker names are represented in bold and the markers commonly identified in DA and QTL analyses are represented as bold underlined. L3, L6, L15 and L35 are the linkage groups from *S. officinarum* whereas S3 is from *S. spontaneum.* The asterix on S3 indicates a segregation distorted marker.

6.5 Conclusions

Molecular breeding *via* marker assisted selection (MAS) is being used as complementary practice in traditional plant breeding methods where improvement of a quantitative trait has been difficult or inefficient (Morgante and Salamini, 2003). In this study, we detected several marker-QTL associations for Brix, sucrose and pol using an interspecific *S. officinarum* x *S. spontaneum* cross. The important aspect of this study is

consistency of QTL detected for Brix and sucrose across different seasons within a crop-year and across different crop-years. Such markers (or QTL) have a potential role in sugarcane breeding via marker assisted selection. Another notable aspect of this research was the identification of QTLs having a positive effect on sucrose accumulation from the wild *S. spontaneum* and negative QTLs from *S. officinarum*. The positive markers from these *Saccharum* species would have immediate role in introgression breeding which would ultimately lead to widening of the sugarcane genetic base. The potential of early harvesting could only be achieved with the deployment of clones that accumulate economic levels of sugar early in the season. Marker assisted selection can assist in this endeavor if markers unique to early season accumulation of sucrose were found. This study uncovered a few markers that could serve as a primer for MAS for high, early sucrose varieties. Several SRAP and TRAP markers were also detected affecting the sugar traits in sugarcane which may have direct application in MAS. The DA results when compared with those from QTL analysis found that a good proportion of the DA-identified markers were either similar or in the vicinity of QTLs implying the potential of DA in MAS. DA also identified several markers that were not linked on the linkage map such that they could not be detected by the QTL analysis. We are currently trying to increase the marker density of *Saccharum* species genetic linkage maps which would serve as anchor map for the Louisiana sugarcane breeding program.

6.6 References

Aitken K. S., P. A. Jackson and C. L. McIntyre. 2005. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor Appl Genet 110:789–801

Aitken, K. S., P. A. Jackson and C. L. McIntyre. 2006. Quantitative trait loci identified for sugar related traits in a sugarcane (*Saccharum* spp.) cultivar × *Saccharum officinarum* population. Theor Appl Genet. 112: 1306-1317

Aljanabi, S. M., Y. Parmessur, H. Kross, S. Dhayan, S. Saumtally, K. Ramdoyal, L. J. C. Autrey and A. Dookun-Saumtally. 2007. Identification of a major quantitative trait locus (QTL) for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. Mol. Breed. 19:1-14

Aluko, G. K. 2003. Genetic mapping of agronomic traits from the interspecific cross *O. sativa* (L.) and *O. glaberrima* (Steud). Ph.D thesis submitted to Louisiana State University, Baton Rouge, LA.

Alwala, S., A. Suman, J.A. Arro, J.C. Veremis, and C.A. Kimbeng. 2006. Target Region Amplification Polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. Crop Sci. 46:448–455

Alwala, S., C. A. Kimbeng, W. P. Williams and M. S. Kang. 2007. Molecular markers associated with resistance to *Aspergillus flavus* in maize: QTL and Discriminant analyses (Manuscript submitted).

Arro, J. A. 2005. Genetic diversity among sugarcane clones using Target Region Amplification Polymorphism (TRAP) markers and pedigree relationships. Masters Thesis submitted to Louisiana State University, Baton Rouge, LA.

Asnaghi, C., D. Roques, S. Ruffel, C. Kaye, J.Y. Hoarau, H. Telismart, J.C. Girard, L.M. Raboin, A.M. Risterucci, L. Grivet, and A. D'Hont. 2004. Targeted mapping of a sugarcane rust resistance gene (*Bru1*) using bulked segregant analysis and AFLP markers Theor. Appl. Genet. 108:759–764

Barret, B. A and K. K. Kidwell. 1998. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci 38:1261–1271

Bernacchi, D., T. Beck-Bunn, Y. Eshed, J. Lopez, V. Petiard, J. Uhlig, D. Zamir, S. Tanksley (2004). Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. Theor. Appl. Genet 97: 381-397

Bouchez, A., F. Hospital, M. Causse, A. Gallais and A. Charcosset. 2002. Marker assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics 162: 1945-1959

Breaux, R.D. 1987. Some breeding strategies with bi-parental and polycrosses. Pages 71-85. In: Proceedings of Copersucar Intern. Soc. Sugar Cane Technol. Sugarcane Breeding Workshop. May/June 1987

Capdevielle, F.M., G. K. Aluko, M. Balzarini and J.H. Oard. 2000. Application of molecular markers and discriminant analysis to identify rice lines with contrasting phenotypes for agronomic traits. In: G.S. Khush, D.S. Brar and B. Hardy (Eds.), Proceedings of the Fourth International Rice Genetics Symposium, International Rice Research Institute, 216 pp., Los Banos, Philippines (abstr.).

Churchill, G. A and R. W. Doerge. 1994. Empirical Threshold Values for Quantitative Trait Mapping. Genetics 138: 963-971

Cox, M. C., D. M. Hogarth and P. B. Hansen .1994. Breeding and selection for high early season sugar content in a sugarcane (*Saccharum* spp. hybrids) improvement program. Aust J Agric Res 45:1569–1575

Cox, M.C. 1999. Improving early-season CCS in a sugarcane breeding program. Proc. 11th Aust. Plant Breed. Conf. 2: 100–101.

Cox, M.C., D. M. Hogarth and R. T. Mullins. 1990. Clonal evaluation of early sugar content. Proc. Aust. Soc. Sugar Cane Technol. 12: 90–98

Cruz-Castillo, J. G., S. Ganeshanandam, B. R. MacKay, G. S. Lawes, C. R. O. Lawoko and D. J. Woolley. 1994. Applications of canonical discriminant analysis in horticultural research. Hortscience 29: 1115-1119

D'Hont, A., L. Grivet, P. Feldmann, S. Rao, N. Berding and J. C. Glaszmann. 1996. Characterisation of the double-genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Mol. Gen. Genet. 250: 405-413

Daugrois, J. H., L. Grivet, D. Roques, J. Y. Hoarau, H. Lombard, J. C. Glaszmann and A. D'Hont. 1996. A putative major gene for rust resistance linked with a RFLP markers in sugarcane cultivar 'R570'. Theor. Appl. Genet. 92:1059–1064

Deren, C.W. 1995. Genetic base of U.S. mainland sugarcane. Crop Sci. 35:1195-1199

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Ann Eugen 7:179–188

Gravois, K.A., B. L. Legendre and K. P. Bischoff. 2002. Cultivar and crop effects of sugarcane bull shoots on sugarcane yield in Louisiana. J. Am. Soc. Sugarcane Technol. 22: 42–52.

Gravois, K. A and S. B. Milligan. 1992. Genetic relationships between fiber and sugarcane yield components. Crop Sci. 32: 62-66

Hoarau, J. Y, B. Offmann, A. D'Hont, A. M. Risterucci, D. Roques, J. C. Glaszmann and L. Grivet. 2001. Genetic dissection of a modern sugarcane cultivar (Saccharum spp.). 1. Genome mapping with AFLP markers. Theor Appl Genet 103:84–97

Hoarau, J. Y., L. Grivet, B. Offmann, L. M. Raboin, J. P. Diorflar, J. Payet, M. Hellmann, A. D'Hont and J. C. Glaszmann. 2002. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTLs for yield components. Theor Appl Genet. 105:1027–1037

Jackson, P. A and D. M. Hogarth. 1992. Genotype _ environment interactions in sugarcane. I. Patterns of response across locations and crop-years in North Queensland. Aust. J. Agric. Res. 43, 1447–1459

Jackson, P. A and T. E. Morgan. 2003. Early stage selection for commercial cane sugar (CCS) in sugarcane clones: effects of time of sampling and irrigation. Aust J Agric Res 54:389–396

Jackson, P. A. 2005. Breeding for improved sugar content in sugarcane. Field Crops Res. 92: 277-290

Jannoo, N., L. Grivet, M. Seguin, F. Paulet, R. Domaingue, P. S. Rao, A. Dookun, A. D'Hont and J. C. Glaszmann. 1999. Molecular investigation of genetic base of sugarcane cultivars. Theor. Appl. Genet. 99: 171-184

Jansen, R. C and P. Stam. 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136: 1447–1455

Kang, M. S., J. D. Miller and P. Y. P Tai. 1983. Genetic and phenotypic path analysis and heritability in sugarcane. Crop Sci. 23: 643-647

Kang, M. S., J. D. Miller, P. Y. P. Tai, J. L. Dean and B. Glaz. 1987. Implications of confounding genotype X year and genotype X crop effects in sugarcane. Field Crops Res. 15: 349-355

Klecka, W. R. 1980. Discriminant analysis. Quantitative applications in the social sciences series, no.19. Sage, Thousand Oaks Lachenbruch PA (1975) Discriminant analysis. Hafner, New York

Kulwal, P. L., R. Singh, H. S. Balyan and P. K. Gupta. 2004. Genetic basis of pre-harvest sprouting tolerance using single-locus and two locus QTL analyses in bread wheat. Genomics 4: 94–101

Lecomte, L., A. Gautier, A.Luciani, P. Duffé, F. Hospital, M. Buret and M. Causse. 2004. Recent advances in molecular breeding: The example of tomato breeding for flavor traits. Acta Hort 637: 231-242

Legendre, B. L. 2001. Sugarcane production handbook. Louisiana State University Agricultural Center Publication # 2859

Li, G., and C.F. Quiros. 2001. Sequence related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: Its application to mapping and gene tagging in Brassica. Theor Appl Genet. 103:455–461

Li, Z., Pinson, S. R. M., Stansel, J. W and A. H. Paterson. 1998. Genetic dissection of the source–sink relationship affecting fecundity and yield in rice (*Oryza sativa* L.). Mol Breed 4: 419–426

Liu, Z. H., J. A. Anderson, J. Hu, T. L. Friesen, J. B. Rasmussen and J. D. Faris. 2005. A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci Theor. Appl. Genet. 111: 782-794

Martin, F. A. 1996. Survey of germplasm needs for *Saccharum* species in the United States. Crop Germplasm committee meeting.

Mcharo, M., D. R. LaBonte, C. Clark, M. Hoy and J. H. Oard. 2005. Molecular marker variability for southern root-knot nematode resistance in sweetpotato. Euphytica 144: 125–132

Mcharo, M., D. R. LaBonte, J. H. Oard, S. J. Kays and W. L. McLaurin. 2004. Linking quantitative traits with AFLP markers in sweetpotato using discriminant analysis. Acta Hort. 637: 285-293

Miklas, P. N., J. Hu, N. J. Grünwald and K. M. Larsen. 2006**.** Potential application of TRAP (Targeted Region Amplified Polymorphism) markers for mapping and tagging disease resistance traits in common bean Crop Sci 2006 46: 910-916

Milligan, S. B., K. A. Gravois and F. A. Martin. 1992. Inheritance of sugarcane ratooning ability and the relationship of younger crop traits to older crop traits. Crop Sci. 36: 45-50

Milligan, S. B., K. A. Gravois, K. P. Bischoff and F. A. Martin. 1990. Crop effects on broad sense heritabilities and genetic variances of sugarcane yield components. Crop Sci. 30: 344-349

Ming, R., S. C. Liu, P. H. Moore, J. E. Irvine and A. H. Paterson. 2001. QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res 11:2075–2084

Morgante, M and F. Salamini. 2003. From plant genomics to breeding practice. Curr. Opin. Biotechnol 14:214–219

Raboin, L-M., K. M. Oliveira, L. Lecunff, H. Telismart, D. Roques, M. Butterfield, J. Y. Hoarau and A. D'Hont. 2006. Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene Theor. Appl. Genet. 112: 1382-1391

Reffay, N., P. Jackson, K. Aitken, J Hoarau, A. D'Hont, P. Besse and C. L. McIntyre. 2005. Characterization of genome regions incorporated from an important wild relative into Australian sugarcane. Mol. Breed. 15: 367-381

Sills, G.R., W. C. Bridges, S. M. Al-janabi and B. W. S. Sobral. 1995. Genetic analysis of agronomic traits in a cross between sugarcane (*Saccharum officinarum* L.) and its presumed progenitor (S. robustum Brandes & Jesw. Ex Grassl). Mol Breed 1:355–363

Sreenivasan, T.V., B.S. Ahloowalia, and D.J. Heinz. 1987. Cytogenetics. p. 211–253. In D.J. Heinz (ed.) Sugarcane improvement through breeding. Elsevier, Amsterdam

Tanksley, S. D., and S.R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063–1066

Tanksley, S. D., S. Grandillo, T. M. Fulton, D. Zamir, T. Eshed, V. Petiard, J. Lopez and T. Beck-Bunn. 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium* Theor. Appl. Genet 92: 213-224

Utz, H. F. 2001. PLABSTAT: A Computer Program for Statistical Analysis of Plant Breeding Experiments, Version 2P. Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart.

Utz, H. F., and A. E. Melchinger. 1996. PLABQTL: a program for composite interval mapping of QTL. J. Quant. Trait Loci 2: 1

Visscher, P. M., C. S. Haley and R. Thompson. 1996. Marker-assisted introgression in backcross breeding programs. Genetics 144:1923-1932

Wang, T., S. S. Xu, M. O. Harris, J. Hu, L. Liu and X. Cai. 2006. Genetic characterization and molecular mapping of Hessian fly resistance genes derived from *Aegilops tauschii* in synthetic wheat. Theor Appl Genet. 113: 611-618\

Zeng, Z.-B. 1994. Precision mapping of quantitative trait loci. Genetics 136: 1457-1468

Zhang, N., Y. Xu, M. Akash, S. McCouch and J. H. Oard. 2005. Identification of candidate markers associated with agronomic traits in rice using discriminant analysis. Theor Appl Genet. 110: 721-729

CHAPTER 7 SUMMARY

7.1 Evaluation of TRAP Markers

Target Region Amplification Polymorphism (TRAP) is a fairly new PCR-based molecular marker technique which uses gene-based information for primer design. Thirty genotypes from the genera *Saccharum*, *Miscanthus*, and *Erianthus* were used in the study. Among the genus *Saccharum* were the species, *S. officinarum* L., *S. barberi* Jesw., *S. sinense* Roxb., *S. spontaneum* L., *S. robustum* Brandes and Jeswiet ex Grassl, cultivars, cultivar-derived mutants and interspecific hybrids between *S. officinarum and S. spontaneum*. Both the cluster and PCoA analyses placed the *Erianthus spp.* and *Miscanthus spp* genotypes distinctly from each other and from the *Saccharum* species, thus, supporting their taxonomic classification as separate genera. Genotypes of the low sucrose and cold tolerant species, *S. spontaneum,* formed one distinct group while the rest of the *Saccharum* species formed one inter-related cluster with no distinct sub-groups. Sequence analysis of TRAP bands derived from a *S. spontaneum* genotype revealed homology with known gene sequences from other grass species including *Sorghum*. A BLASTn search using the homologous sequences from *Sorghum* matched with the *S. officinarum* GenBank accession from which the fixed TRAP primer was designed. These results ratify TRAP as a potentially useful marker technique for genetic diversity studies in sugarcane.

7.2 Comparison of TRAP with AFLP and Coefficient of Parentage

The choice of crossing parents is the most crucial step in any crop improvement program. A better understanding of genetic diversity among the available parental genotypes could help the breeder to make better crosses. In this study, TRAP, AFLP, and pedigree data were used to estimate genetic similarity (GS) among nine sugarcane

genotypes often used as parents. Twelve TRAP primer combinations produced a total of 444 bands, out of which 242 (55 %) were polymorphic, whereas 28 AFLP primer combinations produced a total 1325 bands out of which 686 (53 %) were polymorphic. TRAP-based GS estimates ranged from 0.67 to 0.87 with a mean of 0.75, while AFLP-based estimates ranged from 0.72 to 0.84 with a mean of 0.76. The COP-based GS estimates ranged from 0.03 to 0.36 with a mean of 0.12. The dendrograms were constructed using the unweighted pair-group method with arithmetic mean (UPGMA). Although no distinct pattern was observed in the COP dendrogram, the TRAP dendrogram was better explained by the pedigree records. The AFLP dendrogram showed some distinct cluster patterns. The associations between TRAP-COP (r = 0.41) and AFLP-COP (r = 0.42) were moderate, whereas the TRAP-AFLP (r = 0.14) association was low. Our results indicate that all the three methods estimate different aspects of GS. Therefore, based on the objectives of the research, some combination of TRAP, AFLP and COP would be a better choice in making decisions of which parents to cross in a crop improvement program.

7.3 Predicting the Cross Progeny Performance

Estimating genetic similarity or divergence among parental genotypes in a breeding program could not only aids in predicting genetic parameters but also guides in planning crosses for hybrid/cultivar development. The experimental materials used in this study comprised of five families derived from bi-parental crosses involving nine parents and each family consisted of thirty clones. All the five families were evaluated in the field during the years 2004 (plant cane crop) and 2006 (second ratoon crop). Phenotypic data for stalk height (SH), stalk count (SC), stalk diameter (SD), cane yield (CY) and theoretical recoverable sugars (TRS) were recorded in both years. No data was

recorded in 2005 due to hurricanes. Family means, genetic variances (σ^2_g), average mid parental heterosis (AMPH), percent heterotic clones per family (PHC) and mid parental (MP) estimates were determined for all the traits combined across both years. No significant differences were observed for family means (combined across two years), however, significant differences were observed for clones within each family. The family B (derived from genetically similar and adapted parents) produced high σ^2_g for most of the traits followed by family D (adapted x adapted cross derived from genetically divergent parents). From this study, family means seems to be a fair indicator of breeding potential of cross as positive correlations were observed with AMPH and PHC especially for TRS which is an economic trait of interest in sugarcane breeding. Likewise, positive correlations were observed between σ^2_g and AMPH and PHC for TRS. Negative correlations were predominantly observed between AFLP-GS and most of the genetic parameters except MP values and on the other hand, positive correlations were observed in the case of TRAP-GS. The correlations between *f*-GS and genetic parameters were similar to those observed for TRAP-GS. It appears each method of estimating GS gave reliable information independent of each other, therefore, we recommend a combination of two or more measures to accurately predict the genetic parameters or for use in designing the crosses.

7.4 Genetic Linkage Mapping

Framework genetic linkage maps of two progenitor species of cultivated sugarcane, *Saccharum officinarum* 'Louisiana Striped' (2n=80) and *S. spontaneum* 'SES 147B' (*2n*=64) were constructed using amplified fragment length polymorphism (AFLP), sequence related amplified polymorphism (SRAP) and target region amplification polymorphism (TRAP) markers. The mapping population comprised of 100 F_1 progeny

derived from the *Saccharum* interspecific cross. A total of 344 polymorphic markers were generated from the female (*S. officinarum*) parent out of which 247 (72%) were single dose (segregating in a 1:1 ratio) and 33 (10%) were double dose (segregating in a 3.3:1 ratio) markers. Sixty-four (18%) markers deviated from Mendelian segregation ratios. Likewise, in the *S. spontaneum* genome, out of a total of 306 markers, 221 (72%) were single dose and 43 (14%) were double dose markers whereas 42 markers (14%) deviated from Mendelian segregation ratios. Linkage maps with Kosambi map distances were constructed using a LOD score of > 5.0 and a recombination threshold of 0.45. In *Saccharum officinarum*, 146 markers were linked to form 49 linkage groups (LG) spanning 1732 cM whereas in *S. spontaneum,* 121 markers were linked to form 45 LG spanning 1491 cM. The estimated genome size of *S. officinarum* 'Louisiana Striped' was 2487 cM whereas that of *S. spontaneum* 'SES 147B' was 3232 cM. The genome covered was found to be 69% in *S. officinarum* and 46% in *S. spontaneum*. The *S. officinarum* parent 'Louisiana Striped' behaved like an auto-allopolyploid whereas *S. spontaneum* 'SES 147B' behaved like a true autopolyploid. Although a huge amount of disparity appear to exist between the two genomes, the existence of simple duplex markers, which are heterozygous in both parents and segregate 3:1 in the progeny, in the population tends to affirm that pairing and recombination can occur between the two genomes. The study also revealed that, compared to AFLP, the SRAP and TRAP markers appear less effective as tools for rapidly generating a large number of genome-wide markers for linkage mapping.

7.5 Quantitative Trait Loci (QTL) Analysis

In this study, we report identification of markers associated with sugar related QTLs using a F_1 population derived from a *S. officinarum* 'Louisiana Striped' x *S.*

spontaneum 'SES 147B' cross. Genetic linkage maps of *S. officinarum* and *S. spontaneum* were produced using AFLP, SRAP and TRAP marker techniques. The mapping population was evaluated for Brix (B), sucrose (S) and pol (P) at the early (E) and late (L) plant growing season in 2004 and 2005 (04EB, 04LB, 04LS, 04LP, 05EB, 05ES and 05EP). Composite interval mapping (CIM) QTL analysis was performed to identify marker-trait associations for all the seven traits using the PLABQTL software. In *S. officinarum*, combined across all the traits a total of 50 QTLs were observed with LOD scores ranging from 2.51 to 7.64. The phenotypic variation (adjusted R^2) explained by all QTLs per trait ranged from 15.9% (04LP) to 47.8% (04EB). In *S. spontaneum* a total of 26 QTLs were observed with LOD scores ranging from 2.69 to 7.51 and the phenotypic variation ranging from 6.5% (04LP) to 43.5% (04LB). Thirty four digenic interactions (*i*QTL) were observed in *S. officinarum* whereas only four were observed in *S. spontaneum*. Most of the QTLs observed in *S. officinarum* were positive as expected, although a few negative QTLs were also observed. On the contrary in *S. spontaneum*, in addition to negative QTLs, interestingly several positive QTLs were also found. Two positive marker-QTLs (AF1286 on LG L6 and AF475 on LG L8) in *S. officinarum* and one marker-QTL (PM0983 on LG S5) in *S. spontaneum* were repeatedly observed for Brix and sucrose both at the early and late seasons and across both years. The positive markers observed from *S. officinarum* and *S. spontaneum* would serve as a starting point in marker assisted selection (MAS) to be implemented in introgression breeding and the negative markers could be monitor to purge the deleterious alleles. Our study also indicates that SRAP and TRAP markers could potentially be used for QTL tagging and could be integrated onto already existing linkage maps.

In this study we also evaluated the potential of discriminant analysis (DA) to identify markers associated with Brix, sucrose and pol in *Saccharum* species. The mapping population was categorized into three groups (low, medium and high Brix, sucrose and pol). Discriminant analysis (DA) was performed using PROC DISCRIM and STEPDISC options of SAS. The DA-identified the markers associated with all traits under study with high levels of correct classification of the population. A few of the DA-identified markers were either similar to those identified by the traditional QTL analysis or localized to the same genomic regions as identified by QTL analysis in both parents. In addition, several markers which were not identified by QTL analysis were identified by DA approach. The results from our study indicate that DA could be used as viable approach to identify markers.

APPENDIX

PERMISSION LETTERS

To: "Sreedhar Alwala" <salwal1@lsu.edu>
CC: "Fran Katz" <fkatz@agronomy.org>
Subject: RE: Permission Requested
Date: Monday, June 18, 2007 1:16:36 PM

Dear Sreedhar,

I'm sure there will be no problem. However, I will pass this to the Publications Director just in case.

Craig Roberts
Professor, Division of Plant Sciences
State Forage Specialist
Editor-in-Chief, Crop Science Society of America
108 Waters Hall
University of Missouri
Columbia, MO 65211
http://www.plantsci.missouri.edu/roberts

-----Original Message-----
From: Sreedhar Alwala [mailto:salwal1@lsu.edu]
Sent: Monday, June 18, 2007 10:30 AM
To: Roberts, Craig
Subject: Permission Requested

Dear Dr. Roberts,
Editor-in-Chief of Crop Science

I am Sreedhar Alwala from Louisiana State University.

I kindly request you to grant me the permission to use the below
Mentioned published research paper as 'part' of my dissertation work.

Alwala, S., A. Suman, J.A. Arro, J.C. Veremis, and C.A. Kimbeng. 2006. Target Region Amplification Polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. Crop Sci. 46:448-455.

Thank you and kind regards
--Sreedhar.

To: "'Sreedhar Alwala'" <salwal1@lsu.edu>
CC:
Subject: RE: Permission Requested
Date: Tuesday, June 19, 2007 1:31:48 PM

Hello Sreedhar,

You the permission of the Journal of the American Society of Sugar Cane Technologists to use the journal article,
"Alwala, S., Kimbeng, C. A., Gravois, K. A., and Bischoff, K. P. 2006. TRAP, a new tool for sugarcane breeding: comparison with AFLP and coefficient of parentage. J of Amer. Soc. of Sugar Cane Tech. 26: 62-86", as you deem appropriate for your dissertation.

Congratulations on earning your Ph.D.

Sincerely,

Scott Milligan
Journal of the American Society of Sugar Cane Technologists, Managing Editor
684 Turtle Lane
LaBelle, FL 33935
scottbmilligan@earthlink.net

-----Original Message-----
From: Sreedhar Alwala [mailto:salwal1@lsu.edu]
Sent: Monday, June 18, 2007 1:42 AM
To: scottbmilligan@earthlink.net
Subject: Permission Requested

Dear Dr. Milligan,
Greetings

I kindly request you to grant me the permission to use below mentioned published research paper as a part of my dissertation work.

Alwala, S., Kimbeng, C. A., Gravois, K. A., and Bischoff, K. P. 2006. TRAP, a new tool for sugarcane breeding: comparison with AFLP and coefficient of parentage. J of Amer. Soc. of Sugar Cane Tech. 26: 62-86

Looking forward to hearing from you

Thank you and kind regards
--Sreedhar.

VITA

Sreedhar Alwala was born in 1975 to Mr and Mrs. Rajalakshman Alwala in Secunderabad City of Telangana, India. He had pursued his high school education in the same city and finished his bachelor's degree in Acharya N. G. Ranga Agricultural University, Rajendranagar campus, Hyderabad, India. He then moved to Tamil Nadu Agricultural University to finish his master's degree in plant breeding and genetics. After a short period of stay at University of Saskatchewan, Canada, he moved to Louisiana State University, Baton Rouge, Louisiana, USA to pursue doctoral studies in plant breeding and genetics under the supervision of Dr. Collins Kimbeng. While finishing his doctoral studies, he was married to Ms. Deepthi Kondoju.

Sreedhar Alwala is a receipient of LSU Graduate School Dissertation fellowship award and Gerald O. Mott meritorious student award from Crop Science Society of America. He is currently a member of Crop Science Society of America, American Society of Plant Biologists and Sigma Xi Honor society. He has attended various conferences and presented his research work.