Theor Appl Genet (2006) 113:331–343
DOI 10.1007/s00122-006-0300-8

ORIGINAL PAPER

Characterisation of single nucleotide polymorphisms in sugarcane ESTs

**Giovanni M. Cordeiro · Frances Eliott ·
C. Lynne McIntyre · Rosanne E. Casu ·
Robert J. Henry**

Received: 12 September 2005 / Accepted: 21 April 2006 / Published online: 20 May 2006

Abstract Commercial sugarcane cultivars (*Saccharum* spp. hybrids) are both polyploid and aneuploid with chromosome numbers in excess of 100; these chromosomes can be assigned to 8 homology groups. To determine the utility of single nucleotide polymorphisms (SNPs) as a means of improving our understanding of the complex sugarcane genome, we developed markers to a suite of SNPs identified in a list of sugarcane ESTs. Analysis of 69 EST contigs showed a median of 9 SNPs per EST and an average of 1 SNP per 50 bp of coding sequence. The quantitative presence of each base at 58 SNP loci within 19 contiguous sequence sets was accurately and reliably determined for 9 sugarcane genotypes, including both commercial cultivars and ancestral species, through the use of quantitative light emission technology in pyrophosphate sequencing. Across the 9 genotypes tested, 47 SNP loci were polymorphic and 11 monomorphic. Base frequency at individual SNP loci was found to vary approximately twofold between Australian sugarcane cultivars and more widely between cultivars and wild species. Base quantity was shown to segregate as expected in the IJ76-514 × Q165 sugarcane mapping population, indicating that SNPs that occur on one or two sugarcane chromosomes have the potential to be mapped. The use of SNP base frequencies from five of the developed markers was able to clearly distinguish all genotypes in the population. The use of SNP base frequencies from a further six markers within an EST contig was able to help establish the likely copy number of the locus in two genotypes tested. This is the first instance of a technology that has been able to provide an insight into the copy number of a specific gene locus in hybrid sugarcane. The identification of specific and numerous haplotypes/alleles present in a genotype by pyrophosphate sequencing or alternative techniques ultimately will provide the basis for identifying associations between specific alleles and phenotype and between allele dosage and phenotype in sugarcane.

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-006-0300-8 and is accessible for authorized users.

Communicated by E. Guiderdoni

G. M. Cordeiro (✉) · F. Eliott · R. J. Henry
Centre for Plant Conservation Genetics,
Southern Cross University, PO Box 157,
Lismore 2480, Australia
e-mail: gcordeir@scu.edu.au

C. L. McIntyre · R. E. Casu
CSIRO Plant Industry, Queensland Bioscience Precinct,
306 Carmody Road, St Lucia, QLD 4067, Australia

G. M. Cordeiro · F. Eliott · C. L. McIntyre · R. E. Casu ·
R. J. Henry CRC for Sugar Industry Innovation through
Biotechnology, St Lucia, Australia

Introduction

The genome of modern sugarcane cultivars is a complex blend of aneuploidy and polyploidy derived from interspecific hybridisation. Most sugarcane cultivars contain more than 100 chromosomes which can be assigned to 8 homology groups (Aitken et al. 2005; Rossi et al. 2003). Over the past two decades, studies utilising various molecular techniques to unravel the complexity of this important crop species have provided a greater understanding of its complex genetic makeup (D'Hont 1994; Grivet et al. 1996; Ming et al. 2001; Rossi et al. 2003; Sills et al. 1995; Wu et al. 1992). Significant achievements include milestones that demonstrate the use of single (markers present on one chromosome only) and double dose (marker present on two chromosomes) markers for mapping and QTL analysis (Aitken et al. 2004; Hoarau et al. 2002; Ming et al. 2001, 2002) and large-scale EST sequencing projects by SUCEST (Vettore et al. 2001), SASRI (Carson et al. 2000), UGA (S.R. Schulze, H.M. Ma, J. Meizhu Yang, J.E. Bowers, E. Mirkov, A.H. Paterson, unpublished) and CSIRO (Casu et al. 2004). Whilst genome mapping in sugarcane has made significant progress, it continues at a pace far slower than with many other agricultural crops such as barley (Ramsay et al. 2000), rice (Delseny et al. 2001) and sorghum (Mullet et al. 2002). A suite of tools is available for detailed molecular analysis and has been and is being widely used to provide an insight into the genomes of predominantly diploid species. A major challenge, however, is to apply these new technologies to understand the complex sugarcane genome.

Single nucleotide polymorphisms (SNPs) are being identified and rapidly mapped to provide a rich source of genetic information with potential for allowing a greater insight into understanding the genic complexity of many organisms. Sequence base substitutions have been well characterised since the advent of sequencing technology in 1977 and, indirectly, SNPs and indels have been the basis of DNA-based genetic markers such as restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs) and RAPDs amongst others. SNPs are present in high frequency in any genome, are amenable to high-throughput analysis and have the ability to reveal hidden polymorphisms where other methods fail (Bhattramakki et al. 2001). The understanding of diseases and genetic variation in human individuals has benefited significantly through the use of SNPs in genomic studies. For example, SNPs in specific human genes have been shown to be associated with the risks of developing cardiovascular disease and susceptibility to Alzheimer's disease (Davignon et al. 1988), with susceptibility for hip osteoarthritis (Mototani et al. 2005), and with increased risk of thrombosis (Bertina et al. 1994; Ridker et al. 1995). In addition, pharmacogenomics uses SNPs to predict responsiveness to drug therapy to provide better response medicines to patients (Pfost et al. 2000). In plants also, a number of studies have been able to link SNPs with phenotypic traits of agronomic interest. These include SNPs identified in a putative betaine aldehyde dehydrogenase 2 gene responsible for the fragrance trait in rice (Bradbury et al. 2005) and SNPs found in the starch synthase IIa gene associated with starch gelatinisation temperature in rice (Waters et al. 2005). These studies highlight the usefulness of SNP markers, demonstrating both the abundance of this marker type and the potential causal association between a single nucleotide alteration and organism phenotype.

In recent years, SNPs have become important as genomic markers, with numerous technical methods developed for their detection (Gut 2001). Unfortunately, the majority of these methods are applicable mainly to diploid genomes where a simple presence/absence of either one or both of the alternative bases would indicate homozygosity or heterozygosity. Sugarcane, however, is a complex polyploid and aneuploid species, containing an estimated 8–14 copies of every chromosome (Aitken et al. 2004; Rossi et al. 2003) with individual plants also being highly heterozygous. Thus, for the sugarcane equivalent of a single copy gene in a diploid, located in a homology group containing 14 chromosomes, up to 14 different alleles could be present, with individual alleles present in varying numbers. Thus, the frequency of a SNP base at a gene locus will be determined by both the number of chromosomes carrying the gene and the number of different alleles and frequency of each allele possessing each SNP base. Hence, any method used to detect SNPs at a particular locus in sugarcane must be able to determine the frequency of each base in different genotypes, rather than only detecting the presence and absence of SNPs. In sugarcane, a SNP may be polymorphic between two genotypes because it is present in one and absent in another or because the SNP base scores differ due to the genotypes possessing different alleles or different numbers of copies of each allele. Such detection systems are generally more complex and expensive than simpler and more common methods used for detecting less complex genomes (Ahmadian et al. 2000; Alderborn et al. 2000; Nurmi et al. 2001; Ross et al. 1998; Storm et al. 2003).

Genetic maps are widely used in plant breeding to identify genomic regions controlling traits of interest. Such information assists in understanding the genetic basis of the target trait, as well as providing DNA markers for use in marker-assisted breeding. In sugarcane, only markers that are present as a single copy in one parent and absent in the second (i.e. single dose [SD] marker) can be incorporated into maps using populations of conventional size (~ 250 progeny) (Wu et al. 1992). In these populations, SD markers segregate in a 1:1 ratio; the approximate map position of double dose (DD markers) can also be deduced. Marker systems such as AFLPs, simple sequence repeats or microsatellites (SSRs) and RFLPs can provide a large number of

SD and DD markers for mapping in sugarcane. Several studies have identified markers linked to quantitative trait loci for sugar- and disease-related traits (Aitken et al. 2005; Daugrois et al. 1996; Hoarau et al. 2002; McIntyre et al. 2005a, b; Ming et al. 2001; Rossi et al. 2003). However, with the exception of a limited number of resistance gene analogues (RGAs) and candidate gene/ESTs mapped as RFLP markers (McIntyre et al. 2005a, b; Ming et al. 2001; Rossi et al. 2003), most of the markers in these studies have been anonymous markers. Currently, a limited number of papers describe the potential value of SNPs for development into a marker system for sugarcane. These include a discussion on the ability of SNPs to delineate a set of 64 ESTs into two groups that are likely to represent two gene family members of 6-phosphogluconate dehydrogenase (Grivet et al. 2001); the delineation of 178 ESTs into three paralogous genes to reveal the expression of an Adh2 and two Adh1 genes in sugarcane (Grivet et al. 2003); the use of SNPs for development into co-dominant cleaved amplified polymorphic sequence (CAPS) markers (Quint et al. 2002); and the use of SNPs to map several candidate genes and ESTs (McIntyre et al. 2006). These reports are an indication that the development of SNP markers identified from EST sequences will provide a valuable marker system for mapping candidate genes and for identifying the genetic basis of QTLs of agronomically important traits.

In this study, we have selected a suite of candidate genes, of known and unknown function, which from differential expression studies appear to be up-regulated in maturing cane stem. Using this suite of sugarcane genes, we have investigated the frequency of SNPs in sugarcane ESTs and the use of pyrophosphate sequencing technology to quantify each base at a SNP locus. We demonstrate the utility of this marker system for such applications as genotyping, mapping and determining allele haplotypes of candidate genes.

Materials and methods

The identification of SNPs for development into a marker may be accomplished in a number of ways. In this study, due to the abundance of available sugarcane EST sequences in public databases, it was possible to use the relatively low cost method of in silico SNP discovery.

Selection of candidate genes and sequences

An initial list of 100 candidate genes was selected based on studies (Casu et al. 2003, 2004) involving differentially expressed transcripts from maturing stem of sugarcane. EST sequences from candidate genes were matched with tentative unique contigs (Saccharumtuc) obtained from the clustering of sugarcane ESTs by PlantGDB on 22/01/2004 (http://www.plantgdb.org/), a National Science Foundation (USA) funded project to develop plant species-specific EST databases. ESTs used by PlantGDB in the clustering were obtained from GenBank and clustered using the program CAP3 (Huang et al. 1999). Of the 100 candidate genes, 69 matched with tentative unique contigs.

SNP identification

The SNP identification was carried out using an in-house developed Perl script that colour codes columns with mismatched characters. Putative SNPs selected for marker development were from contigs that comprised a minimum of five ESTs in the alignment and where both preceding and proceeding bases were non-identical to the putative SNP bases. Sufficient bases were required between SNPs for sequencing primer design.

Plant material

Nine sugarcane genotypes were selected for the initial screening of markers (Table 1). The selection included commercial and ancestral cultivars, progeny of a mapping population and a noble cane. Eight are parents of four mapping populations. The ninth genotype, Q124, is a widely cultivated commercial cultivar. In addition, up to 80 progeny of the IJ76-514 × Q165 segregation cross were used to measure base quantities at individual SNP loci in a selected set of candidate genes. DNA from all genotypes was a gift from K. Aitken (CSIRO Plant Industry, Brisbane, Australia).

PCR and primer design

The use of the Pyrosequencer requires: (1) a single-stranded DNA template in which the SNP or SNPs reside; and (2) a SNP detection primer that is complimentary to the template with its 3′ end between 0 and 4 bp from the SNP. Unlike the single base primer extension assay (Syvänen 1999), there is no requirement for the primer to anneal adjacent to the polymorphic nucleotide site.

For ease of design, the SNP detection primer is preferentially designed in the forward direction and complimentary to the template strand. Only the template strand is required in the Pyrosequencer. To separate the template strand, it is biotinylated in a PCR amplifi-

Table 1 Genotypes used in this study

Genotype	Species	Chromosome number (2*n*)	Comment[a]
Q117	*Saccharum* spp. hybrid	~ 107–109	Australian commercial cultivar; parent of segregation cross Q117 × MQ77-340
Q124	*Saccharum* spp. hybrid	Unknown	Australian commercial cultivar
Q162	*Saccharum* spp. hybrid	Unknown	Australian commercial cultivar; parent of BC1 segregation cross Q162 × KQ99-1391
Q165	*Saccharum* spp. hybrid	~ 115	Australian commercial cultivar; parent of segregation cross IJ76-514 × Q165
MQ77-340	*Saccharum* spp. hybrid	Unknown	Elite sugarcane parent; parent of segregation cross Q117 × MQ77-340
Mida	*Saccharum* spp. hybrid	Unknown	Australian commercial variety; parent of BC1 segregation cross KQ99-1410 × Mida
KQ99-1391	*Saccharum* spp. hybrid	Unknown	Progeny of IJ76-514 × Q165 population; parent of BC1 segregation cross Q162 × KQ99-1391
KQ99-1410	*Saccharum* spp. hybrid	Unknown	Progeny of IJ76-514 × Q165 population; parent of BC1 KQ99-1410 × Mida
IJ76-514	*S. officinarum*	80	Noble cane; parent of segregation cross IJ76-514 × Q165

[a] By convention, the parent used as the female in a cross is written first

cation process through the use of a biotin-labelled primer. The biotinylation of primers in a developmental phase is highly costly. Hence, an alternative method utilising a universal biotin-labelled 11-mer for attachment to the template strand was utilised (Pacey-Miller et al. 2003). Briefly, the method involves two sequential PCR amplification reactions. The first PCR amplification uses a desalted forward primer and a reverse primer with an additional 11 base tag sequence 5′-GCCCCCGCCCGNNNNNN...NNNNN-3′. The second PCR amplification step uses the same desalted forward primer and a biotin-labelled 11 base tag, with the sequence 5′-(Biotin)GCCCCCGCCCG-3′, as the reverse primer. This produces PCR products with a biotin-labelled reverse strand.

The criteria for optimal PCR primer design were for the amplification of a template product between 80 and 200 bp, with shorter fragments preferred. PCR primer length was set at 21±3 bp; T_m between 50 and 65°C; GC content between 40 and 60%; and complementarity of 2 bp or less at the 3′-end.

Criteria for the selection of sequencing primers were a primer length between 13 and 19 bp; average T_m of ~ 50°C (range between 43 and 53°C); 3′-end between 0 and 5 bp from the identified SNP; and no 3′-end complementarity greater than 2 bp due to the low running temperature of the Pyrosequencer at 28°C. Where a putative SNP was immediately flanked by a base identical to either SNP base, primer design was abandoned due to the difficulty in partitioning the contribution to the frequency by the actual SNP and that of the surrounding identical bases.

All primers, with the exception of the biotin-labelled oligonucleotide, were designed using the software program, Primer Premier v 5.00 by PREMIER Biosoft International. All biotin-labelled oligonucleotides were synthesised and HPLC purified by Proligo, Singapore. All others were synthesized by Proligo, Lismore, Australia.

The PCR template primers were used to amplify genomic DNA from the genotypes tested. First amplification round PCR was carried out in a total of 25 µl containing 20 ng template DNA, 0.2 µM of forward and reverse primers, 0.2 mM of each dNTP, 1.0 mM $MgCl_2$ and 1 U Platinum Taq polymerase (Invitrogen—Life technologies) in the supplied buffer. Reactions were undertaken on a Perkin Elmer 9700 thermocycler. Cycling conditions were 3 min at 94°C followed by 35 cycles of 10 s at 94°C, 30 s at the appropriate annealing temperature (50–65°C), 30 s at 75°C and a final extension step of 5 min at 75°C. In the second PCR amplification to attach the biotinylated tag sequence, the relevant forward or reverse primer (depending on the sequence direction of template strand desired) was replaced with the biotin tag with amplification being carried out over 40 cycles, with

25 ng DNA and 0.5 mM $MgCl_2$. All other conditions remained the same.

Pyrophosphate sequencing

Pyrophosphate sequencing uses the principal of de novo incorporation of nucleotides (Nyrén et al. 1985; Ronaghi et al. 1996, 1998). Each incorporation event is accompanied by the release of pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide. The release of PPi triggers a cascade of enzymatic reactions that culminates in the production of light that is then detected by a charge-coupled device or CCD camera. This is seen as a peak in a pyrogram™ with each light signal proportional to the number of nucleotides incorporated.

Labelled PCR products were separated for sequencing using the Pyrosequencing™ Vacuum Prep Tool. Three microlitres of Streptavidin Sepharose™ HP (Amersham) was added to 37 μl binding buffer (10 mM Tris–HCl, pH 7.6, 2 M NaCl, 1 mM EDTA, 0.1% Tween 20) and mixed with 20 μl PCR product and 20 μl high-purity water for 10 min at room temperature using an Eppendorf Thermomixer (15 min at 1,400 rpm). The magnetic beads containing the immobilised templates were captured by rare earth magnets and transferred to denaturation solution (0.5 M NaOH) for 5 s, then onto a wash buffer 10 mM Tris–acetate, pH 7.60, for 5 s. The vacuum was then released and the beads released into a PSQ 96 Plate Low containing 45 μl annealing buffer (20 mM Tris–acetate, 2 mM $MgAc_2$, pH 7.6) and 0.3 μM sequencing primer.

The pyrophosphate sequencing reaction is carried out automatically within the PSQ 96 system (Pyrosequencing AB, Uppsala, Sweden) using a SNP reagent kit according to the manufacturer's instructions and the appropriate SNP detection primer. Within the Pyrosequencer, pyrophosphate sequencing is performed in a volume of 50 μl and an operating temperature of 28°C. Pyrosequencer scores are analysed using the Allele Frequency Quantification function in the proprietary software by Pyrosequencing AB.

Results

SNPs identified in ESTs

Of the 100 candidate genes selected from a list of differentially expressed ESTs in maturing cane stem (Casu et al. 2004), 69 were identified in 990 clustered ESTs deposited in the PlantGDB database (http://www.plantgdb.org). Each cluster of sequences comprised between 2 and 68 ESTs with an average of 14.35 ESTs. Each consensus sequence averaged 1,150 bp in length making a total of approximately 79,320 bases of sequence. The number of putative SNPs per cluster of ESTs ranged between 0 and 107, making a total of 1,588 putative SNPs across the 79,320 bases or an average of 23 (median of 9) SNPs per EST sequence or 1 SNP per 50 bp. InDels were not included in this estimate.

Marker development

Contigs comprising five or more sequences (53 in total) were selected for SNP marker development. Taking into consideration the criteria for primer design, 175 SNP detection primers were designed to just over 180 putative SNPs in 19 contigs. This represents just over one-tenth of the total number of putative SNPs identified.

To determine the technical repeatability of the pyrophosphate sequencing system in sugarcane, DNA was isolated from leaf material from two genotypes, Q117 and Q124, collected from six geographical regions within the Australian state of Queensland. These DNA samples were scored using two SNP markers (SuSNP068-T388 and SuSNP077-A2155) developed as part of an initial pilot study (data not shown). For both genotypes and both SNPs, the allele scores were highly repeatable, with a variance between 0.42 and 1.36% and a percentage base composition differing by a maximum of 2.4% (Table 2).

A total of 123 PCR primer pairs were designed to amplify templates corresponding to the candidate genes. As a set of PCR template amplification primers may cover a fragment length encompassing more than

Table 2 Base scores of two genotypes derived from six separate geographical regions

Genotype	Q117 (%:%)	Q124 (%:%)
Marker: SuSNP068-T388		
Location 1	A71.8:G28.2	A73.3:G26.7
Location 2	A74.5:G25.5	Data not available
Location 3	A72.0:G28.0	Data not available
Location 4	A71.5:G28.5	A74.0:G26.0
Location 5	A73.6:G26.4	A74.6:G25.4
Location 6	A72.7:G27.3	Data not available
Variance	±1.36%	±0.42%
Maximum variation	2.1%	1.3%
Marker: SuSNP077-A2155		
Location 1	Data not available	T38.5:C61.5
Location 2	T62.2:C37.8	T36.1:C63.9
Location 3	T59.8:C40.2	T38.2:C61.8
Location 4	T61.3:C38.7	T36.7:C63.3
Location 5	T61.1:C38.9	T37.0:C63.0
Location 6	T61.8:C38.2	T36.5:C63.5
Variance	±0.83%	±0.93%
Maximum variation	2.4%	2.4%

one SNP detection primer, the number of SNP detection primers was greater than the number of PCR primer pairs designed and synthesized. In addition, several SNP detection primers detected more than one SNP locus when these loci were only 2–5 bp apart.

More than 1,500 pyrophosphate sequencing runs were performed to identify potential SNP markers of value. Attrition of both primer types resulting from complete amplification failure, non-specific priming or self-priming, reduced the number of SNP detection primers to 53 and PCR template amplification primers to 45 pairs. As some SNP detection primers detected more than one SNP locus, a total of 58 SNP loci were detected by the 53 detection primers. Primer sequences of both PCR and SNP detection primers are listed in Table S1. Of the 58 SNP loci, 48 were polymorphic and 10 monomorphic across the 9 genotypes tested. These 58 markers fell into 19 contiguous sequence sets (Table 3). Each SNP marker was named according to the tentative unique contig it was derived from and the SNP location on that contig.

Base frequency scores were measured for all 58 SNP loci in the 9 genotypes; as a representation of these data, scores for 5 SNP loci, crcSNP13343-A153, crcSNP14965-T473, crcSNP16213-C1234, crcSNP19357-A503 and crcSNP19357-C869 across the 9 genotypes are presented in Table 4. Each genotype has a unique combination of base scores, illustrating how these markers provide

Table 3 Developed markers in relation to the contiguous sequences they were derived from

Contig number[a]	Classification[b]	Putative function[b]	EST[b]	No. of detection primers	Nature of detected SNPs[c]	Total[d]
1328	Chromatin and DNA metabolism	Histone protein	MCS009G11	1	1 polymorphic	1
3473	No assigned function	Significant but no function assigned	MCSA142C12	1	1 polymorphic	1
5209	No assigned function	Significant but no function assigned	MCSA070G11	1 1	1 polymorphic 1 monomorphic	2
5597	Membrane transport	Lipid-transfer protein	YCS35.072	2	2 polymorphic	2
5876	Primary metabolism	Phenylalanine ammonia-lyase	MCSA034B10	2	1 polymorphic 1 monomorphic	2
5938	Carbohydrate metabolism	Beta-galactosidase	YCS43.096	4	4 polymorphic	4
9123	Primary metabolism	ATP citrate-lyase	YCS39.068	7	7 polymorphic 1 monomorphic	8
9844	No assigned function	Significant but no function assigned	YCS15.038	2	1 polymorphic 1 monomorphic	2
10813	Carbohydrate metabolism	Beta-glucosidase	YCS15.042	2	1 polymorphic 1 monomorphic	2
11026	No assigned function	Significant but no function assigned	MCSA061E12	1	1 polymorphic	1
13343	Primary metabolism	Hydrolase	MCSA035A02	3	3 polymorphic	3
14965	No assigned function	Significant but no function assigned	MCSA141D07	2	3 polymorphic	3
14977	Defence/stress-related proteins	DAHP synthase	MCSA114C05	2	2 polymorphic	2
15943	Primary metabolism	Phenylalanine ammonia-lyase	MCSA042E11	4	4 polymorphic	4
16213	Fibre biosynthesis and degradation	Caffeic acid 3-*O*-methyltransferase	MCSA064G01	5	7 polymorphic 1 monomorphic	8
19357	No assigned function	Significant but no function assigned	MCSA118F07	2	2 polymorphic	2
22783	No assigned function	Significant but no function assigned	MCS006G04	2	1 polymorphic 1 monomorphic	2
23154	Gene expression and RNA metabolism	RNA binding protein	MCSA116D01	4	3 polymorphic 1 monomorphic	4
27819	Cell wall structure or metabolism	Proline-rich protein	YCS43.031	5	3 polymorphic 2 monomorphic	5
				53	48 polymorphic 10 monomorphic	58

[a] Contig number is derived from contigs as assembled by PlantGDB on 22 January 2004. The prefix 'Saccharumtuc' has been omitted

[b] As determined by Casu et al. (2003, 2004)

[c] A marker is considered polymorphic when base scores differ in at least one out of the nine genotypes tested

[d] Each detection primer may detect more than one SNP. Hence the total number of SNPs detected can be greater than the number of primers

unique fingerprints for individual genotypes. Base frequency scores for the remaining markers that scored as polymorphic are presented in Table S2.

Sugarcane lines KQ99-1391 and KQ99-1410 are progeny from a cross between IJ76-514 and Q165. For the five SNPs, the base scores of the two progeny lines (KQ99-1391 and KQ99-1410) are intermediate between the base scores of the two parents (Table 4). Across the eight Australian hybrid genotypes, the base frequencies varied 1.75±0.09 fold, although this difference is closer to twofold when calculated over the full set of markers tested.

Table 4 Fingerprints of the nine genotypes based on five SNP markers

Genotype	Marker	Base score
IJ76-514	crcSNP13343-A153	A13:G87
	crcSNP14965-T473	C0:T100
	crcSNP16213-C1234	C55:T45
	crcSNP19357-A503	A60:G40
	crcSNP19357-C869	C40:T60
KQ99-1391	crcSNP13343-A153	A22:G78
	crcSNP14965-T473	C13:T87
	crcSNP16213-C1234	C81:T19
	crcSNP19357-A503	A55:G45
	crcSNP19357-C869	C45:T55
KQ99-1410	crcSNP13343-A153	A20:G80
	crcSNP14965-T473	C11:T89
	crcSNP16213-C1234	C60:T40
	crcSNP19357-A503	A67:G33
	crcSNP19357-C869	C54:T46
Mida	crcSNP13343-A153	A33:G67
	crcSNP14965-T473	C0:T100
	crcSNP16213-C1234	C70:T30
	crcSNP19357-A503	A50:G50
	crcSNP19357-C869	C79:T21
MQ77-340	crcSNP13343-A153	A33:G67
	crcSNP14965-T473	C18:T82
	crcSNP16213-C1234	C48:T52
	crcSNP19357-A503	A71:G29
	crcSNP19357-C869	C66:T34
Q117	crcSNP13343-A153	A33:G67
	crcSNP14965-T473	C0:T100
	crcSNP16213-C1234	C80:T20
	crcSNP19357-A503	A54:G46
	crcSNP19357-C869	C54:T46
Q124	crcSNP13343-A153	A23:G77
	crcSNP14965-T473	C0:T100
	crcSNP16213-C1234	C50:T50
	crcSNP19357-A503	A50:G50
	crcSNP19357-C869	C79:T21
Q162	crcSNP13343-A153	A18:G82
	crcSNP14965-T473	C0:T100
	crcSNP16213-C1234	C69:T31
	crcSNP19357-A503	A36:G64
	crcSNP19357-C869	C64:T36
Q165	crcSNP13343-A153	A33:G67
	crcSNP14965-T473	C9:T91
	crcSNP16213-C1234	C67:T33
	crcSNP19357-A503	A54:G46
	crcSNP19357-C869	C54:T46

As *Saccharum officinarum* is an octoploid, there are eight copies of each chromosome. Thus, single dose bases at SNP loci in IJ76-514 should be approximately 13:87 or vice versa. As sugarcane varieties such as Q165 have an estimated 12 copies of each chromosome, single dose bases at SNP loci should be approximately 8:92. Similarly, DD bases should be approximately 25:75 or 17:83 for the two parental lines, respectively. On this basis, the number of tentative single, double, triple and multi-dose markers across the parents of the two segregation and two back cross populations are presented in Table 5. The putative dosage for each marker was subjectively determined by its proportional representation in the genome. Due to a lack of information on the exact number of alleles of each gene in a sugarcane genotype, it is impossible to determine the exact dosage without scoring the markers on a large number of progeny from each segregating population. The information does, however, provide an indication of the proportion of SD and DD markers that can ultimately be used for mapping.

Confirmation of selected SD and DD markers identified in the IJ76-514 × Q165 segregation cross was carried out by running the relevant markers across a subset of progeny from the IJ76-514 × Q165 mapping population (Table 6). Single (markers present once in one parental genome) and double (markers present twice in one parent) dose markers segregate in a ratio of 1:1 and 11:3, respectively (da Silva et al. 1993). Each of the segregation ratios of the SD and DD markers was tested against the expected ratios using a χ^2 test for the respective markers. Segregation ratios did not significantly differ from the expected ratios in all three cases, confirming the dose level of the markers in the parents.

Table 7 shows base frequency scores vary across the different SNP loci of the contig Saccharumtuc.5938 providing an indication that the locus is represented by several alleles. To ascertain the likely copy number of alleles at homo(eo)logous loci, the possible ratios for each frequency score were initially determined. When determining the possible ratios, based on the results in Table 2, a variation of up to ±2.4% was allowed for each base frequency score. Hence, in most instances, a base frequency score could represent two or more possible ratios. By using a combination of multiple SNP loci from a homo(eo)logous locus, it becomes possible to determine the likely copy number of the gene based on the recurrence of a common possible copy number across the assayed loci. For example, the base

Table 5 Putative SNP base dosage as identified in the parents of four segregation crosses

Cross	Monomorphic	Multi-dose	Triple dose	Double dose	Single dose	Total markers
Q117 × MQ77-340	16	30	0	5	2	53
IJ76-514 × Q165	13	30	5	6	4	58
Average	15	30	2.5	5.5	3	55.5

Table 6 Putative SD, DD and MD markers tested on a subset of progeny from the IJ76-514 × Q165 segregation cross to confirm dosage

Marker	Dose	SNP	Base proportion (%:%)		Progeny scored	Segregation ratio
			IJ76-514	Q165		
crcSNP14965-T254	DD	T/C	T100:C0	T83:C17	76	25DD:40SD:11ND (65:11)
crcSNP14965-T473	SD	T/C	T100:C0	T91:C9	79	34SD:45ND
crcSNP16213-T1310	SD	T/C	T100:C0	T92:C8	40	16SD:24ND

ND null dose or absent

Table 7 Variation of SNP base frequencies from six SNP loci in contig Saccharumtuc.5938 and the deduced likely copy number of the contig in two sugarcane genotypes

Marker	Base frequency score(%:%)	Possible ratios	Possible copy numbers
Genotype: Mida			
crcSNP5938-A1663	A59.3:G41.7	6:4 or 7:5	10 or 12
crcSNP5938-G1776	T58.0:G42.0	6:4 or 7:5	10 or 12
crcSNP5938-T2026	C33.1:T66.9	3:6 or 4:8	9 or 12
crcSNP5938-C2083	C34.8:G66.2	3:6 or 4:8	9 or 12
crcSNP5938-C2338	C57.0:T43.0	8:6 or 7:5	14 or 12
crcSNP5938-C2377	C59.3:T40.7	6:4 or 7:5	10 or 12
	Likely copy number		12
Genotype: KQ99-1410			
crcSNP5938-A1663	A63.1:G36.9	5:3 or 7:4	8 or 11
crcSNP5938-G1776	T63.5:G36.5	5:3 or 7:4	8 or 11
crcSNP5938-T2026	C28.6:T71.4	3:8	11
crcSNP5938-C2083	C37.1:G62.9	3:5 or 4:7	8 or 11
crcSNP5938-C2338	C62.8:T37.2	3:5 or 4:7	8 or 11
crcSNP5938-C2377	C60.0:T40.0	6:4 or 7:5	10 or 12
	Likely copy number		11

frequency call of A59.3:G41.7 for the genotype Mida has two possible ratios 6:4 or 7:5 that represent possible copy numbers of either 10 or 12. Repeating this deduction across the assayed loci, the possible copy numbers for Mida would range between 9 and 12. The copy number 12 is, however, consistent for each base frequency score across the 6 SNPs, indicating this to be the most likely copy number for this locus. Similarly, the most likely copy number for the locus in the genotype KQ99-1410 is 11.

In addition, the base frequency scores for the genotype Mida indicate the presence of at least three alleles in this genotype. From Table 7, the SNPs at positions 1,663 (A base), 1,776 (T base), 2,338 (C base) and 2,377 (C base) have a similar frequency score of approximately 58% and could be on the same haplotype or allele. SNPs at positions 2,026 (C base) and 2,083 (C base) have a lower base frequency of approximately 33% and could either be on the subset of the higher frequency (i.e. 58%) haplotypes or on the subset of the lower frequency haplotypes (i.e. 41%) of the four SNPs above. The third allele would be the remaining of the base frequencies, i.e. 9% [100% − (58% + 33%)]. Hence, the first allele could have the most common base (58%) of SNPs 1,663, 1,776, 2,338 and 2,377, and the most common base (58%) at SNPs 2,083 and 2,338 giving a possible haplotype of ATTGCC; the second allele could have the least common base (33%) at all 6 SNPs giving the possible haplotype of GGCCTT; and the third allele could have the least common base at SNPs 1,663, 1,776, 2,338 and 2,377 and the most common base (9%) at SNPs 2,026 and 2,083 giving a possible haplotype of GGTGTT. For the genotype KQ99-1410, it can be surmised that at least four alleles are present as SNPs 2,026 and 2,083 occur at different frequencies.

Discussion

The application of SNP marker technology to the polyploid and aneuploid genome of sugarcane presents complications not encountered with diploid genomes. For example, the frequency of a heterozygous SNP locus would be represented by a frequency of 50:50 for each SNP base and 100:0 or 0:100 at a homozygous SNP locus in a diploid organism. Often, there is no need to determine the frequency of individual bases at SNP loci, and detection systems that simply provide a presence/absence result are sufficient to determine the allelic variant present.

In sugarcane, the proportional frequencies of each SNP base will vary depending on the number of alleles of the gene containing the SNP locus. The ability to capture this information accurately across several SNPs within a set of homo(eo)logous alleles can give an indication of the number of allele haplotypes present for a gene and potentially provide the haplotype sequences. This information could have implications for sugarcane breeding. High performing sugarcane lines could be due to the presence of a specific allele(s) present at a gene locus or to a different number of copies of a specific allele at a gene locus or to a combination of both. Knowledge of the sequence underlying each allele haplotype has the potential to allow allele-specific markers to be designed. However, information on allele numbers and frequency are required before these questions can be investigated.

Single nucleotide polymorphisms occur relatively frequently in the genomes of all organisms. Varying frequencies of SNPs per length of DNA sequence have been reported; however, this is highly dependent upon the gene and whether coding or non-coding regions are examined. Studies using genomic DNA have found 1 SNP per 83 bases in the analysis on 8 maize inbred lines (Bhattramakki et al. 2001); in barley, approximately 1 SNP per 27 bases were identified in the intronless Isa gene (Bundock et al. 2004), and a study on the exonic region of the P450 gene family members in barley found approximately 1 SNP per 131 (Bundock et al. 2003). In human, one estimate of nucleotide diversity across a set of 850 full-length coding sequences was placed at approximately 1 in 3,300 bases (Garg et al. 1999), whilst estimates based on whole genome data place the frequency at 1 in 1,000 bases (Sachidanandam et al. 2001; Venter et al. 2001). In sugarcane ESTs, an average of 1 SNP per 50 bases was identified. With each EST contig averaging approximately 1,150 bp, that would yield an average of 23 (or a median of 9) potential SNPs for analysis.

Of the 58 SNP loci scored by the 53 detection primers, 42 loci were polymorphic and 12 monomorphic. During the SNP selection process, an alternate base had to occur at least twice in the EST cluster to avoid selecting SNPs that may be a result of low-quality sequence trace data. In a number of cases where the developed marker has been identified as monomorphic, the alternate base was identified as present in more than two ESTs. For example, in the marker crcSNP10813-G1751, the putative SNP was identified at the ratio of 22G (consensus base):6C (alternate base); and in the marker crcSNP23154-C973, the putative SNP was present at a ratio of 16C:7T. Both these markers were monomorphic across the nine genotypes tested. This is, however, not unexpected as the majority of EST sequences in the public database are derived from genotypes developed in Brazil, whilst the genotypes tested in here are, with the exception of IJ76-514, developed in Australia. Hence, with a different set of genotypes, the ratio of polymorphic to monomorphic markers may differ.

An assessment of the technical repeatability of pyrophosphate sequencing was performed by testing two markers (SuSNP068-T388 and SuSNP077-A2155) on six individuals, each of two genotypes Q117 and Q124, collected from six separate geographical regions in Queensland, Australia. The results showed a variance between 0.77 and 2.6% in pyrophosphate sequencing scores. This variation in scores corresponds closely with the 1.1–3.8% variance observed using the same system in tetraploid potato (Rickert et al. 2002), indicating the level of accuracy achievable with the system. The base ratios were less than 1:2 in both cases.

Germplasm fingerprinting using SNPs

As with other genetic marker systems, SNPs can be used for germplasm fingerprinting and marker-assisted breeding (Bhattramakki et al. 2001) through the judicial selection of a set of informative markers. The United States Department of Agriculture estimates a set of 43 well-selected markers will have sufficient power to uniquely identify all 10 million cattle ever registered with the American Angus Association (Heaton et al. 2002). Whilst the high ploidy level of the sugarcane genome is often seen as a disadvantage in genome analysis, the resulting heterozygosity from the mix of *S. officinarum* and *S. spontaneum* chromosomes

(Bremer 1961) increases heterozygosity, allowing a small number of markers to positively identify a relatively large number of genotypes. In Table 4, it was shown that the base scores from five markers are capable of uniquely identifying each of the nine genotypes tested. It is also possible to select just two of the five markers listed to uniquely identify the nine genotypes.

Sugarcane germplasm is maintained as living nurseries and the accurate identification of clones in collections is a major issue in sugarcane breeding and varietal exchange programs. The accurate fingerprinting of clones with a set of informative SNP markers will help eliminate duplicates in collection, eliminate clone mislabelling and identify diverse germplasm and resources of genes for specific breeding targets. The fingerprinting of clones will also assist in disputes over varieties protected under Plant Breeders Rights.

Markers for mapping

Two markers putatively identified as SD in the IJ76-514 × Q165 parents showed the expected segregation ratio of ~ 1:1 when scored on the progeny of the segregation cross (Table 6). A third marker putatively identified as DD segregated in the expected 11:3 ratio indicating that these markers have the potential for being mapped onto sugarcane maps. However, the robustness of the pyrophosphate sequencing system is challenged by the high ploidy level of the sugarcane genome and the 1:9 or 1:10 ratio of SD markers pushes the limits of the dynamic range of the system, making the calling of these extreme frequencies by the software unreliable. Attempts at mapping the two SD markers using the Pyrosequencer have failed due to this limitation (Rickert et al. 2002; Ronaghi 2001; Ronaghi et al. 1996, 1998) and alternative methods such as mass spectrometry will need to be tested.

Many assay methods are available for SNP detection and include hybridization, enzymatic methods, primer extension, oligonucleotide ligation, 5′-nulclease assay, light emission and mass spectrometry (Gut 2001; Syvänen 2001). Whilst the ability to accurately quantify SNP base frequencies provides additional information about a locus and has the potential to determine allelic haplotypes, it is not necessary for the purposes of gene mapping. Mapping in sugarcane requires the efficient detection of markers that are present as SD or DD. Hence, methods such as eco-tilling, which provides no quantifiable information of each SNP base, can be used for mapping if a SNP can be identified as being single dose (McIntyre et al. 2006).

In silico SNP discovery for the specific purpose of developing markers for mapping does have its limitations. There is little information in aligned EST sequences to indicate which putative SNPs will potentially represent single dose markers in a particular segregation cross and this needs to be verified empirically. Out of the 69 candidate genes assessed, there were 42 polymorphic SNPs across the 9 genotypes tested. Just four markers were identified as putative SD in the parents of the IJ76-514 × Q165 segregation cross and the average number of putative SD markers in the parents of all four segregation crosses was 3.25 (Table 5). Even without testing the putative SD markers on the progeny of the segregation crosses for confirmation, this represents between 4.7 and 9.5% of all polymorphic markers developed. Approximately 71% of polymorphic AFLPs are SD and 75.6% of polymorphic SSR bands (Aitken et al. 2005). At least 50 of 55 RGA RFLPs generated at least 1 SD marker and could be mapped (Rossi et al. 2003). Whilst the number of SD markers that can be generated appears small compared to other marker systems, it should be emphasised that not all potential SNPs were developed into markers due to the restraints placed on primer design.

The discovery process for SD markers is better approached through the application of the tilling method (Till et al. 2004) where multiple SNP loci can be identified on a selected fragment of DNA. Applying this method on a sample progeny population from a segregation cross of interest will present potential segregation ratios of each SNP. The SNP can then either be developed into a marker for use on an alternative detection system such as the Sequenom or the remaining population ‘eco-tilled’ (Comai et al. 2004) to map single dose markers. We have shown in a separate publication (McIntyre et al. 2006) that SD SNPs in sugarcane can be mapped through the eco-tilling process.

As a marker for association mapping

Modern sugarcane cultivars are derived from a small number of crosses between an estimated 19 *S. officinarum* plants and 5 *S. spontaneum* plants. Linkage disequilibrium in sugarcane has been found to extend several cM in a low-resolution RFLP study (Jannoo et al. 1999).

In theory, the association of SNP variations with either the presence or absence of different phenotypes among individuals or among individuals from different populations appears straightforward. This simplistic view does not account for the majority of base polymorphisms that do not result in any amino acid change. Determining the haplotypes is more important for predicting individual phenotypes than are the underlying SNPs. Determining haplotypes also allows the ability to infer the evolutionary history of a DNA region (Templeton et al. 1988; Tishkoff et al. 1998). However, difficulties are encountered in determining SNP haplotypes when inbred or homozygous individuals are not available (Rafalski 2002) as is usually the case with sugarcane.

However, with the pyrophosphate sequencing technology, the ability to determine SNP base frequencies provides the means to determine the likely copy number of homo(eo)logous loci (Table 7). Where chromosome counts have been performed for a genotype, this information can be used to support the inference of the most likely copy number of homo(eo)logous loci. Knowledge of the number of homo(eo)logous loci will assist in the deduction of the allelic composition of the locus in any particular sugarcane genotype.

The ability to determine haplotypes opens possibilities in unravelling the complexities of the sugarcane genome. By defining haplotypes in parents of crosses, it may be possible to deduce their segregation in progeny; or it could be used to determine allele dosage and composition in any particular genotype to determine phenotypic performance.

A number of computational methods exist to resolve haplotypes from combinations of SNP ratios from a single locus. These methods utilise such algorithms as the expectation-maximization (EM) algorithm (Clayton 1990; Excoffier et al. 1995; Fallin and Shork 2000; Hawley et al. 1995); Bayesian approaches (Niu et al. 2002; Stephens and Donnelly 2003; Stephens et al. 2001) and algorithms based on parsimony (Gusfield 2001; Lancia et al. 2004). Without exception, these methods have all been designed with the diploid human genome in mind. The challenge now is to adapt these methods to suit the ploidy level of sugarcane.

Summary

Through mining of the public EST databases, we have determined that SNPs occur at high frequency in the sugarcane genome. We have demonstrated that there is considerable variation in base compositions at SNP loci between sugarcane genotypes which can be used to fingerprint and identify individual genotypes. We have also demonstrated that this variation in base composition segregates as expected in progeny of mapping populations. The determination of individual allele haplotypes within sugarcane by combining the information on base composition at multiple SNPs within a gene still requires further investigation. Single dose SNPs appear to occur in sugarcane ESTs at a lower frequency than SD markers generated using other methods; however, alternative methods of discovery may help resolve this issue for ESTs that code for agronomically desirable traits. Ongoing research includes the investigation of alternative SNP detection methods, such as eco-tilling for SNP discovery and mapping of single dose markers in a member of the sucrose phosphate synthase gene family (McIntyre et al. 2006).

Acknowledgements This research was undertaken with funding from the Cooperative Research Centre for Sugar Industry Innovation through Biotechnology.

References

Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyrén P, Uhlén M, Lundeberg J (2000) Single-nucleotide polymorphism analysis by pyrosequencing. Anal Biochem 280:103–110

Aitken K, Jackson P, Piperidis G, McIntyre L (2004) QTL identified for yield components in a cross between a sugarcane (*Saccharum* spp.) cultivar Q165^A and a *S. officinarum* clone IJ76-514. In: Proceedings for the 4th international crop science congress, Brisbane, Australia, 26 September–1 October 2004. http://www.cropscience.org.au

Aitken KS, Jackson PA, McIntyre CL (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

Alderborn A, Kristofferson A, Hammerling U (2000) Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. Genome Res 10:1249–1258

Bertina RM, Koelemann BPC, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, van der Velden PA, Reitsma PA (1994) Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature 369:64–67

Bhattramakki D, Rafalski A (2001) Discovery and application of single nucleotide polymorphism markers in plants. In: Henry RJ (ed) Plant genotyping: the DNA fingerprinting of plants. CAB International, Lismore, pp 179–191

Bradbury LMT, Fitzgerald TL, Henry RJ, Jin QS, Waters DLE (2005) The gene for fragrance in rice. Plant Biotechnol J 3:363–370

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

Bundock PC, Henry RJ (2004) Single nucleotide polymorphism, haplotype diversity and recombination in the Isa gene of barley. Theor Appl Genet 109:543–551

Bundock PC, Christopher JT, Eggler P, Ablett G, Henry RJ, Holton TA (2003) Single nucleotide polymorphisms in cytochrome P450 genes from barley. Theor Appl Genet 106:676–682

Carson DL, Botha FC (2000) Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci 40:1769–1779

Casu RE, Grof CPL, Rae AL, McIntyre CL, Dimmock CM, Manners JM (2003) Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol Biol 52:371–386

Casu RE, Dimmock CM, Chapman SC, Grof CPL, McIntyre CL, Bonnett GD, Manners JM (2004) Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. Plant Mol Biol 54:503–517

Clayton DG (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol 7:111–122

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S (2004) Efficient discovery of DNA polymorphisms in natural populations by ecotilling. Plant J 37:778–786

Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H, Glaszmann JC, Dhont A (1996) A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570'. Theor Appl Genet 92:1059–1064

Davignon J, Gregg RE, Sing CF (1988) Apolipoprotein E polymorphism and arteriosclerosis. Arteriosclerosis 8:1–21

Delseny M, Salses J, Cooke R, Sallaud C, Regad F, Lagoda P, Guiderdoni E, Ventelon M, Brugidou C, Ghesquière A (2001) Rice genomics: present and future. Plant Physiol Biochem 39:323–334

D'Hont A (1994) A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. Genome 37:222–230

Excoffier L, Slatkin M (1995) Maximum-likelihood-estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927

Fallin D, Shork NJ (2000) Accuracy of haplotype frequency estimation for biallelic loci, via the EM algorithm for unphased diploid genotype data. Am J Hum Genet 67:947–959

Garg K, Green P, Nickerson DA (1999) Identification of candidate coding region single nucleotide polymorphisms in 65 human genes using assembled expressed sequence tags. Genome Res 9:1087–1092

Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann J-C (1996) RFLP mapping in a highly polyploid and aneuploid interspecific hybrid. Genetics 142:987–1000

Grivet L, Glaszmann JC, Arruda P (2001) Sequence polymorphism from EST data in sugarcane: a fine analysis of 6-phosphogluconate dehydrogenase genes. Genet Mol Biol 24:161–167

Grivet L, Glaszmann JC, Vincentz M, da Silva F, Arruda P (2003) ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. Theor Appl Genet 106:190–197

Gusfield D (2001) Inference of haplotypes from samples of diploid populations: complexity and algorithms. J Comput Biol 8:305–323

Gut IG (2001) Automation in genotyping of single nucleotide polymorphisms. Hum Mutat 17:475–492

Hawley ME, Kidd KK (1995) HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. J Hered 86:409–411

Heaton MP, Harhay GP, Bennett GL, Stone RT, Grosse WM, Casas E, Keele JW, Smith TPL, Chitko-McKown CG, Laegreid WW (2002) Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. Mamm Genome 13:272–281

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

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–887

Jannoo N, Grivet L, Seguiin M, Paulet F, Domaingue R, Rao PS, Dookun A, D'Hont A, Glaszmann JC (1999) Molecular investigation of the genetic base of sugarcane cultivars. Theor Appl Genet 99:171–184

Lancia G, Pinotti MC, Rizzi R (2004) Haplotyping populations by pure parsimony: complexity of exact and approximation algorithms. INFORMS J Comput 16:348–359

McIntyre CL, Casu RE, Drenth J, Knight D, Whan VA, Croft BJ, Jordan DR, Manners JM (2005a) Resistance gene analogues in sugarcane and sorghum and their association with quantitative trait loci for rust resistance. Genome 48:391–400

McIntyre CL, Whan VA, Croft B, Magarey R, Smith GR (2005b) Identification and validation of molecular markers associated with pachymetra root rot and brown rust resistance in sugarcane using map- and association-based approaches. Mol Breed 16:151–161

McIntyre CL, Jackson M, Cordeiro G, Amouyal O, Eliott F, Henry RJ, RE Casu, Hermann S, Aitken KS, Bonnett GD (2006) The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. Mol Breed (in press)

Ming R, Liu S-C, Moore PH, Irvine JE, Paterson AH (2001) QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res 11:2075–2084

Ming R, Del Monte TA, Hernandez E, Moore PH, Irvine JE, Paterson AH (2002) Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. Genome 45:794–803

Mototani H, Mabuchi A, Saito S, Fujioka M, Iida A, Takatori Y, Kotani A, Kubo T, Nakamura K, Sekine A, Murakami Y, Tsunoda T, Notoya K, Nakamura Y, Ikegawa S (2005) A functional single nucleotide polymorphism in the core promoter region of CALM1 is associated with hip osteoarthritis in Japanese. Hum Mol Genet 14:1009–1017

Mullet JE, Klein RR, Klein PE (2002) *Sorghum bicolor*—an important species for comparative grass genomics and a source of beneficial genes for agriculture. Curr Opin Plant Biol 5:118–121

Niu T, Qin ZS, Xu X, Liu JS (2002) Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. Am J Hum Genet 70:157–169

Nurmi J, Kiviniemi M, Kujanpaa M, Sjoroos M, Ilonen J, Lovgren T (2001) High-throughput genetic analysis using time-resolved fluorometry and closed-tube detection. Anal Biochem 299:211–217

Nyrén P, Lundin A (1985) Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Anal Biochem 151:504–509

Pacey-Miller T, Henry R (2003) SNP detection in plants using a single stranded pyrosequencing protocol with a universal biotinylated primer. Anal Biochem 317:165–170

Pfost DR, Boyce-Jacino MT, Grant DM (2000) A SNPshot: pharmacogenetics and the future of drug therapy. TIBTECH 18:334–338

Quint M, Mihaljevic R, Dussle CM, Xu ML, Melchinger AE, Lubberstedt T (2002) Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize. Theor Appl Genet 105:355–363

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Ramsay L, Macaulay M, degli Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997–2005

Rickert AM, Premstaller A, Gebhardt C, Oefner PJ (2002) Genotyping of SNPs in a polyploid genome by pyrosequencing (TM). Biotechniques 32:592–603

Ridker PM, Hennekens CH, Lindpainter K, Stampfer MJ, Eisenberg PR, Miletich JP (1995) Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. N Engl J Med 332:912–917

Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. J Chromatogr B 739:345–355

Ronaghi M, Karamohamed D, Petterson B, Uhlén M, Nyrén P (1996) Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 242:84–89

Ronaghi M, Uhlén M, Nyrén P (1998) Real-time pyrophosphate detection for DNA sequencing. Science 281:363–365

Ross P, Hall L, Smirnow I, Haff L (1998) High levelmultiplex genotyping by MALDI-TOF mass spectrometry. Nat Biotechnol 16:1347–1351

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys M-A, D'Hont A (2003) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. Mol Gen Genet 269:406–419

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD et al (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933

Sills G, Bridges W, Al-Janabi S, Sobral BWS (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

da Silva J, Sorrells ME, Burnquist WL, Tanksley SD (1993) RFLP linkage map and genome analysis of *S. spontaneum*. Genome 36:782–791

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162–1169

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

Storm N, Darnhofer-Patel B, van den Boom D, Rodi CP (2003) MALDI-TOF mass spectrometry-based SNP genotyping. Methods Mol Biol 212:241–262

Syvänen A-C (1999) From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum Mutat 13:1–10

Syvänen A-C (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2:930–942

Templeton AR, Sing CF, Kessling A, Humphries S (1988) A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. Genetics 120:1145–1154

Till BJ, Curtner C, Comai L, Henikoff S (2004) Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res 32:2632–2641

Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, Bonne-Tamir B, Kidd JR, Pafstis AJ, Jenkins T, Kidd KK (1998) A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. Am J Hum Genet 62:1389–1402

Venter JC, Adams MD, Myers EW (2001) The sequence of the human genome. Science 291:1304–1351

Vettore AL, da Silva FR, Kemper EL, Arruda P (2001) The libraries that made SUCEST. Genet Mol Biol 24:1–7

Waters DLE, Henry RJ, Reinke RF, Fitzgerald MA (2005) Gelatinisation temperature of rice explained by polymorphisms in starch synthase. Plant Biotechnol J (In press)

Wu K, Burnquist W, Sorrels M, Tew TL, Moore P, Tanksley S (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor Appl Genetics 83:294–300