Theor Appl Genet (2009) 118:327–338
DOI 10.1007/s00122-008-0902-4

ORIGINAL PAPER

Informative genomic microsatellite markers for efficient genotyping applications in sugarcane

Swarup K. Parida · Sanjay K. Kalia · Sunita Kaul · Vivek Dalal · G. Hemaprabha · Athiappan Selvi · Awadhesh Pandit · Archana Singh · Kishor Gaikwad · Tilak R. Sharma · Prem Shankar Srivastava · Nagendra K. Singh · Trilochan Mohapatra

Received: 21 April 2008 / Accepted: 20 September 2008 / Published online: 23 October 2008
© Springer-Verlag 2008

Abstract Genomic microsatellite markers are capable of revealing high degree of polymorphism. Sugarcane (*Saccharum* sp.), having a complex polyploid genome requires more number of such informative markers for various applications in genetics and breeding. With the objective of generating a large set of microsatellite markers designated as Sugarcane Enriched Genomic MicroSatellite (SEGMS), 6,318 clones from genomic libraries of two hybrid sugarcane cultivars enriched with 18 different microsatellite repeat-motifs were sequenced to generate 4.16 Mb high-quality sequences. Microsatellites were identified in 1,261 of the 5,742 non-redundant clones that accounted for 22% enrichment of the libraries. Retro-transposon association was observed for 23.1% of the identified microsatellites. The utility of the microsatellite containing genomic sequences were demonstrated by higher primer designing potential (90%) and PCR amplification efficiency (87.4%). A total of 1,315 markers including 567 class I microsatellite markers were designed and placed in the public domain for unrestricted use. The level of polymorphism detected by these markers among sugarcane species, genera, and varieties was 88.6%, while cross-transferability rate was 93.2% within *Saccharum* complex and 25% to cereals. Cloning and sequencing of size variant amplicons revealed that the variation in the number of repeat-units was the main source of SEGMS fragment length polymorphism. High level of polymorphism and wide range of genetic diversity (0.16–0.82 with an average of 0.44) assayed with the SEGMS markers suggested their usefulness in various genotyping applications in sugarcane.

Communicated by A. Kilian.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0902-4) contains supplementary material, which is available to authorized users.

S. K. Parida · S. K. Kalia · S. Kaul · V. Dalal · A. Pandit · A. Singh · K. Gaikwad · T. R. Sharma · N. K. Singh · T. Mohapatra (✉)
National Research Centre on Plant Biotechnology, Indian Agricultural Research Institute, New Delhi 110012, India
e-mail: tm@nrcpb.org

G. Hemaprabha · A. Selvi
Sugarcane Breeding Institute, Coimbatore, Tamil Nadu 641007, India

S. K. Parida · P. S. Srivastava
Department of Biotechnology, Jamia Hamdard University, New Delhi 110062, India

Introduction

Microsatellites or simple sequence repeats (SSRs) are arranged in tandem repeats of one to six nucleotide long DNA motifs dispersed throughout the eukaryotic genomes. The unique sequences flanking the microsatellite motifs are used to design primers for locus specific amplification. High degree of allelic variation revealed by microsatellite markers results from variation in number of repeat-motifs at a locus caused by replication slippage and/or unequal crossing-over during meiosis (Goldstein and Schlotterer 1999). Microsatellite markers have gained considerable importance in plant genetics and breeding owing to their many desirable genetic attributes including hypervariability, wide genomic distribution, co-dominant inheritance, reproducibility, multi-allelic nature, and chromosome-specific location. These markers are amenable to high-throughput genotyping and are thus suitable for paternity determination, construction of high density genome maps,

mapping of useful genes, marker-assisted selection, and for establishing genetic and evolutionary relationships. A large number of microsatellite markers are now available for completely sequenced plant genomes, namely rice (IRGSP 2005) and *Arabidopsis thaliana* (*Arabidopsis Genome Initiative* 2000). Publicly available expressed sequence tag (EST) and unigene sequences, which correspond to the expressed component of the genome have also been used to design such markers (Parida et al. 2006). Advancement in enrichment techniques and selection of microsatellite containing clones from genomic library has led to the development of genomic microsatellite markers in many plant species including maize (Sharopova et al. 2002), rice (Chen et al. 1997), wheat (Pestsova et al. 2000), and barley (Liu et al. 1996).

The genome of modern sugarcane cultivars is a complex blend of aneuploidy and polyploidy derived from the inter-specific hybridization involving different *Saccharum* species particularly, *S. officinarum* and *S. spontaneum*. Use of an efficient molecular marker system is essential for sugarcane genome for understanding the genetic and taxonomic complexity, and broadening the genetic base of sugarcane cultivars, thereby improving sugar yield and its stabilization against abiotic and biotic stresses. The desirable attributes of microsatellites as DNA markers have encouraged the development (Cordeiro et al. 1999) and utilization of microsatellite markers for many applications in sugarcane genetics and breeding (Rossi et al. 2003; Aitken et al. 2005). Microsatellite markers have also been designed from the publicly available EST databases of sugarcane (Pinto et al. 2004). Being derived from the conserved expressed component of the genome, the EST derived microsatellite markers have shown high degree of cross-transferability, but low level of polymorphism among *Saccharum* species clones and related genera (Cordeiro et al. 2001). In contrast, the microsatellite markers developed from the genomic sequences are reported to be more polymorphic than the EST-derived markers in the *Saccharum* complex (Pinto et al. 2006). A limited number of (about 200) genomic microsatellite markers (International Sugarcane Microsatellite Consortium, ISMC; http://www.scu.edu.au/research/cpcg) till date have been developed in sugarcane from the microsatellite enriched genomic library (Cordeiro et al. 2000) and evaluated for their utility in genome analysis (Pinto et al. 2004; Aitken et al. 2005). Sugarcane with a large polyploid genome requires more such highly informative genomic microsatellite markers for various applications in genetics, genomics and breeding.

The objectives of the present study were to develop a larger set of genomic microsatellite markers, determine their frequency and relative distribution and evaluate their cross-transferability, polymorphic potential, and efficiency in assessment of molecular genetic diversity in sugarcane.

Materials and methods

Construction of genomic library enriched for microsatellites

Two genomic libraries, one of high sugar containing and red-rot resistant popular Indian sugarcane hybrid cultivar Co 7201 enriched for 12 different microsatellite repeat-motifs, namely $(CA)_{16}$, $(GA)_{14}$, $(CAA)_8$, $(AAC)_9$, $(CAC)_9$, $(AGA)_{10}$, $(ACA)_{10}$, $(CAT)_{10}$, $(TTC)_{10}$, $(GAT)_{10}$, $(CTT)_{10}$ and $(GATA)_8$ and another of high sugar containing commercial sugarcane hybrid cultivar Co 86011 enriched for six different microsatellite repeat-motifs, namely $(CGG)_{10}$, $(GCA)_{10}$, $(AAAG)_8$, $(CGGC)_8$, $(CCCCT)_7$ and $(GTCCCG)_6$ were constructed. The total genomic DNA was isolated from fresh leaves and used for constructing genomic library. Minor modifications to an existing microsatellite enrichment procedure (Edwards et al. 1996) enabling optimization for complex plant genomes (Cordeiro et al. 1999) were carried out. Twenty microgram genomic DNA was nebulized and size fractionated to generate DNA fragments in the desired size range (0.5–1 kb). The DNA fragments were purified from gel, end polished and ligated with 1 μg of 21 mer adaptor (CTCTTGCTTAGATCTGGACTA). The ligated fragments were amplified by PCR and passed through streptavidin coated paramagnetic beads (New England Biolabs, Inc., NEB, USA) containing 5′ end biotinylated synthetic oligonuleotide microsatellite repeat-motif probes to select the genomic DNA fragments containing microsatellites. The eluted single stranded enriched DNA fragments were washed two times (5 min per wash) in 6× SSC + 0.1% SDS at 65°C followed by two washes in 6× SSC at room temperature, PCR amplified using the adaptor primers, ligated with the pGEM-T Easy vector (Promega, USA) and transformed into competent DH5α *Escherichia coli* strain by electroporation (BIO-RAD, USA). The transformed cells were screened for blue/white colony and the white positive colonies were selected followed by confirmation for the presence of insert by restriction using *Eco*RI. A total of 6,318 microsatellite enriched genomic clones were picked up from the two genomic libraries and archived in 96-well microtitre culture plates containing 150 μl of LB freezing medium, incubated overnight at 37°C for growth and finally stored in −80°C deep freezer.

Sequencing of microsatellite enriched genomic clones

The genomic clones of the enriched libraries were grown overnight at 37°C in 5 ml LB medium containing 100 μg/ml ampicillin and the plasmid DNA was isolated using R.E.A.L. prep plasmid kit (Qiagen, USA) and sequenced directly with M13 universal primers in both forward and

reverse directions using the capillary-based automated MegaBACE 4000 sequencer (Amersham Biosciences, USA). The trace files were base called and checked for quality using *phred* and assembled using *phrap* software tools. Sequences containing at least 100 continuous nucleotides with a *phred* score greater than 20 were clustered by *phrap* with a minimum consensus phrap score of 80. The assembled contigs were viewed and edited using *consed*. The high-quality sequences thus obtained were used further for mining and characterizing the microsatellites.

Microsatellite mining and primer design

The high quality sequences of the unique genomic clones were searched for microsatellites as described earlier (Parida et al. 2006). The identified microsatellites were designated as Sugarcane Enriched Genomic MicroSatellites (SEGMS), characterized as perfect (monomers to hexamers) and compound (non-interrupting and interrupting) repeat-motifs, and grouped into class I (≥20 nucleotides) and class II (12–20 nucleotides) types. Their relative frequency and distribution in the genomic sequences were estimated. Primers (forward and reverse) were designed in a batch module manner from the flanking sequences of the identified microsatellite motifs employing the microsatellite primer discovery tool (http://hornbill.cspp.latrobe.edu.au/cgi-binpub/indexssr.pl) based on the criteria described earlier (Parida et al. 2006) to develop SEGMS markers. The SEGMS repeat-motifs containing genomic sequences were BLAST searched (http://www.ncbi.nlm.nih.gov/blast.html) against the nucleotide *nr* database and the matching (*E* values of ≤1e-15 and bit score of ≥100) sugarcane genomic sequences were annotated for determining their similarity with the repetitive transposable elements.

Evaluation of amplification efficiency and polymorphic potential

We designed 270 primer-pairs and studied their amplification efficiency using the template DNA of the sugarcane variety Co 7201 from which one genomic library was constructed. The PCR conditions particularly the annealing temperature (varying from 50 to 60°C) for each primer was standardized. Forty-four of these were used to evaluate their polymorphic potential using 21 genotypes belonging to five sugarcane species, three related genera, eight commercial Indian varieties, and five cereal species (Table S1 in electronic supplementary material). The amplified products were resolved in 10% polyacrylamide gel using 0.5× TBE buffer, run at 200 V between 2 and 3 h depending on the size of the expected PCR product, and visualized under UV light following staining with GelStar (CAMBREX Bioscience, USA). The band sizing of the amplicons generated by the SEGMS markers was determined as against 50 bp DNA ladder. The polymorphic information content (PIC) for each SEGMS marker was calculated using the expression, $\text{PIC} = 1 - \sum P_{ij}^2$ (Anderson et al. 1993), where P_{ij} is the frequency of the *j*th allele for *i*th locus summed across all alleles for the locus. Cluster analysis among the *Saccharum* species, varieties and related genera was based on Jaccard's similarity coefficient (Jaccard 1908) by using the unweighted pair group method analysis (UPGMA) and SAHN clustering algorithm in the NTSYS version 2.02e (Applied Biostatistics, Inc.) software package. The confidence limits of UPGMA based dendrogram was determined by bootstrap analysis. Five hundred bootstrap replicates were computed and bootstrap of 50% majority rule consensus tree was constructed using the bootstrap procedure of the WinBoot software program (Yap and Nelson 1996). The potential of 44 SEGMS markers for showing cross-transferability, polymorphism and molecular diversity, and establishing genetic relationships in *Saccharum* species, related genera, varieties, and five cereals was compared with that observed for the 21 unigene derived microsatellite (UGMS) markers designed for sugarcane (Unpublished).

Cloning and sequencing of SEGMS length variants

Amplicons obtained for three SEGMS markers, namely SEGMS33, SEGMS36, and SEGMS122 with $(TGT)_{11}$, $(GA)_{15}$, and $(AC)_{12}$ repeat-motifs, respectively, were eluted from polyacrylamide gel, purified, cloned, and sequenced as described above. Ten random positive clones containing inserts of individual PCR amplicons obtained from each of three SEGMS markers amplified in eight commercial Indian sugarcane varieties including Co 7201 were sequenced in both forward and reverse directions. The high-quality sequences generated after *phred* and *phrap* analysis for all the ten clones of individual variety were assembled. The consensus sequences were derived for each variety and compared among themselves and with the original sequence from which the primers were designed using CLUSTALW multiple sequence alignment tool employing BIOEDIT software (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

Results

Isolation and characterization of microsatellites from the enriched genomic libraries of sugarcane

Two enriched genomic libraries of sugarcane containing 6,318 genomic clones with an average insert size of 659 bp

(varied from 0.5 to 1 kb) were sequenced in both forward and reverse directions (12,636 sequencing reactions) generating 4.16 Mb high-quality sequences after *phred* and *phrap* analysis. Of these, 576 (9.1%) clones (0.66 Mb sequences) were found redundant and thus not considered further (Table S2). Hence, the marker design was based on 3.5 Mb high-quality sequences from 5,742 non-redundant clones. A total of 1,261 clones were identified to have microsatellite repeat-motifs (1,077 perfect and 184 compound excluding mononucleotides) accounting for overall 22% enrichment of the genomic libraries for microsatellites (Table S3). Most (169, 91.8% of the total microsatellites) of the compound microsatellites were interrupting types and the remaining (15, 8.2%) were non-interrupting types. The frequency of mononucleotides was 15% of total perfect microsatellites identified. The mononucleotides showed a strong bias (76.7%) toward C/T repeat-motifs, while the rest contained A/G repeat-motifs. A majority (97.8%) of the mononucleotide repeats were 10–29 bases long and the remaining extended up to 51 bases (C_{51}).

Trinucleotide repeat-motifs were the most abundant class of microsatellite which accounted for 48.5% (621) of all perfect microsatellites identified (Table S4). Trinucleotide motifs were followed by dinucleotide (30.4%, 389), tetranucleotides (13.6%, 174), pentanucleotides (6.5%, 84), and hexanucleotides (0.9%, 11) repeat-motifs. Among the trinucleotide repeats, the motifs $(ACA)_n$ (108, 17.4% of total trinucleotide repeats) were most abundant (Table S4) followed by $(CAA)_n$ (78, 12.5%), and $(AAC)_n$ (57, 9.2%). Among the dinucleotide repeats, CA motifs (80, 20.6% of total dinucleotide repeats) were the most common followed by GA (53, 13.6%) and TA (32, 8.2%). The most abundant tetranucleotide, pentanucleotide, and hexanucleotide repeat-motifs were TCTA (13.2%), GTGTT (13%), and TTCTTT (27.2%), respectively (Table S4). The dinucleotide repeats were the longest with a maximum motif length of 60 nucleotides. The expansion of trinucleotide repeat-motifs was up to 49 times, while those of tetranucleotide, pentanucleotide, and hexanucleotide repeats were up to 12, 13, and 11 times, respectively. A set of 457 longer perfect and 166 compound class I microsatellite containing clones were identified. The density of perfect and compound class I repeat-motifs was one in every 7.7 and 21.1 kb of genomic sequences whereas its proportion was 35.7 and 90.2% of the total microsatellites identified, respectively (Table S4). The class I dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeat motifs were 11.4, 55.6, 22.3, 8.3, and 2.4% of the total perfect class I microsatellites, respectively (Fig. 1a, b). In silico analysis revealed that 230 (18.2%) of the 1,261 microsatellite motif containing enriched genomic sequences showed significant nucleotide level homology (*E* value of ≤1e-141 and bit score of

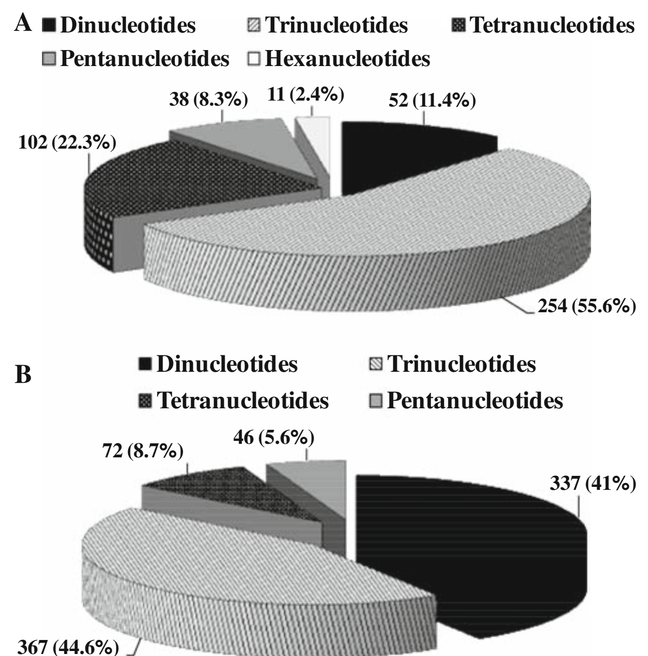

Fig. 1 Comparative distribution of different repeat-motifs under long hypervariable class I (**a**) and potentially variable class II (**b**) microsatellites developed from the enriched genomic library of sugarcane. Trinucleotide was the most abundant repeat motif in both class I (254, 55.6%) and class II (367, 44.6%) category, which was followed by tetranucleotide motifs (102, 22.3%) under class I and dinucleotide motifs (337, 41%) under class II

≥500) to known transposable elements like *Copia*/*Ty1* and *Gypsy*/*Ty3* groups of long terminal repeat (LTR) retrotransposons (Table S5) as well as the non-LTR retrotransposons (62, 5%). Alignment of individual retroelements with the microsatellite containing genomic sequences indicated their presence in the vicinity of microsatellite repeat-motif regions (Fig. S1 in electronic supplementary material).

Development of SEGMS markers and their PCR amplification efficiency

The primers flanking the microsatellite repeat-motifs could be designed for 1,315 (90%) of the 1,463 microsatellite positive clones identified. The primer sequences along with Tm values and product sizes for 1,168 perfect SEGMS markers including 420 class I types are given in the Table S6 A and B. Besides, the primer sequences for 147 compound class I SEGMS markers containing either ten non-interrupting (6.8%) or 137 (93.2%) interrupting types were designed and provided in the Table S7. More than one microsatellite (varying from two to four) either with similar or different perfect repeat-motifs interrupted by ≥100 nucleotides were identified in 102 sugarcane genomic sequences. Each of such multiple repeat-motifs in a

genomic sequence was considered as individual SEGMS for primer designing and thus a total of 1,168 SEGMS markers were designed in this study from 1,077 genomic sequences. All the 1,077 perfect microsatellite containing genomic sequences of sugarcane were submitted to NCBI GenBank (Accession numbers from FI129988 to FI131064) for unrestricted use. To determine amplification efficiency of the SEGMS markers, 270 markers designed from the flanking sequences of class I and class II microsatellite repeat-motif types were chosen and used in PCR amplification. Two hundred and fifty-seven (95.2%) of these gave successful amplification and remaining did not work in any of the 21 genotypes including the one from which the library was constructed. Two-hundred and thirty-six (91.8%) of the 257 amplified SEGMS markers produced fragments of expected size in metaphor agarose gel.

Cross-transferability and polymorphic potential of SEGMS markers

Forty-four SEGMS markers including 28 class I and 16 class II repeat-motif types were used to understand their cross-transferability potential. However, successful amplification of either similar or varying sized fragments obtained with 41 markers in *Saccharum* species and related genera and 11 markers in five cereal species suggested 93.2 and 25% transferability of the SEGMS markers designed from a hybrid sugarcane cultivar to the members of *Saccharum* complex and cereals, respectively. In contrast, with the genic microsatellite markers, the cross-transferability rate was higher (95.2%) to both *Saccharum* complex and cereals. The efficiency of markers to detect polymorphism was studied in a set of 21 genotypes belonging to five sugarcane species, three related genera, eight varieties, and five cereal species (Table 1). Forty-one (93.2%) markers were polymorphic (mean PIC of 0.75) among these genotypes (Fig. 2a), of which 11 (100%, mean PIC of 0.78) amplified dinucleotide repeat-motifs, 28 (90.3%, PIC of 0.70) trinucleotide motifs, and two (100%, PIC of 0.80) tetranucleotide motifs. All the 28 (100%, mean PIC of 0.78) class I markers and 13 (81.2%, mean PIC of 0.73) of the 16 class II markers showed polymorphism among the *Saccharum* species, genera, varieties, and cereals. Thirty-nine (88.6%) of the 41 markers showed polymorphism (mean PIC of 0.76) among the sugarcane species and related genera, whereas 33 (75%) detected polymorphism (mean PIC of 0.79) among the eight sugarcane varieties. A total of 388 fragments were amplified by 41 polymorphic SEGMS markers in the 21 genotypes used in this study. The number of fragments amplified by the polymorphic SEGMS markers varied from three to fifteen with an average of nine per marker in sugarcane.

The potential of 44 SEGMS markers was compared with that of 21 microsatellite markers derived from the unigene sequences of sugarcane to detect polymorphism in the same set of 21 genotypes. The genic microsatellite markers detected lower level of polymorphism (38%) with one to four alleles per locus (Fig. 2b) giving an average PIC of 0.52 in contrast to SEGMS markers. The extent of inter-varietal polymorphism (19%, mean PIC of 0.51) as detected by genic microsatellite markers was much lower (Table 2) as compared to the SEGMS markers.

Molecular basis of the SEGMS fragment length polymorphism

For determining the pattern of SEGMS fragment length variation, the size variant amplicons for the markers SEGMS33, SEGMS36 and SEGMS122 from eight sugarcane varieties were cloned and sequenced. High quality sequence alignment revealed that the size variant amplicons contained the expected microsatellite motif sequences, but variable number of repeat-units with conserved primer binding sites. For instance, the sequence analysis of length variant amplicons at SEGMS33 locus had the expected $(TGT)_n$ motif with the repeat number being either 11 or 15 in different sugarcane varieties (Fig. 3). It thus revealed that the presence of variable number of repeat-units in different amplicons was the major source of SEGMS fragment length polymorphism in sugarcane.

Genetic relationships among the sugarcane genotypes as revealed by SEGMS markers

The SEGMS markers revealed a broader range of pair-wise genetic similarity among 16 genotypes belonging to species, related genera, and commercial varieties of sugarcane that varied from 0.16 to 0.82 with an average of 0.44 as compared to those estimated for the UGMS markers (range 0.50–0.80 with average of 0.67). The similarity among the eight sugarcane varieties varied from 0.57 (CoS 8436 and Co 8021) to 0.80 (Co 8021 and Co 8371) with an average of 0.73 in case of UGMS markers while for SEGMS markers, that ranged from 0.39 (Co 1148 and Co 419) to 0.82 (Co 8021 and Co 8371) with an average of 0.53. Both the marker types revealed higher similarity of the varieties with the sugarcane species than the related genera. The genetic relationships among the genotypes as depicted in dendrograms (Fig. 4, Fig. S2) revealed similar cluster pattern supported by high bootstrap values with the two marker types. All the clones of five *Saccharum* species were included in a major cluster (I), while the commercial sugarcane varieties remained together in a different cluster (II). The three related genera, which were highly divergent

Table 1 The SEGMS markers used for evaluating their cross-transferability and polymorphic potential among 21 genotypes belonging to five cereal species, *Saccharum* complex and varieties

Sl. no.	SEGMS markers	Repeat motifs	Forward primer sequences (5′–3′)	Reverse primer sequences (5′–3′)	Annealing temperature (*T*°C)	Size (bp) range of fragments amplified	Polymorphic (P)/ monomorphic (M)	Range of fragments amplified	PIC value
1	SEGMS2	$(AAC)_{13}$	CGCTATCCTAATAGCCTTCTAA	AAGTCTCTGCTATGTTACGGTT	55	100–1,000	P	6–12	0.78
2	SEGMS7	$(TTC)_{26}$	AGTCTGGCTAAGATCTGGACT	CTCCTATGTTTGAGTTGCATTA	56	100–850	P	5–14	0.80
3	SEGMS10	$(GT)_{25}$	CGTTCGTAGGCAAGAGTGACAG	AGTTAGCAAAAGGCAAAGCACT	55	50–650	P	6–13	0.79
4	SEGMS20	$(TCT)_{11}$	TAGTCCGGCTTAGATCTGGACT	GGAGAATGCTGAAGAATACAAT	55	100–900	P	7–14	0.81
5	SEGMS21	$(ACA)_{17}$	GGCCGTGTGATGCTCCGGCCGC	AGGTACTCCATATGCCAAATAC	55	100–800	P	5–14	0.80
6	SEGMS33	$(TGT)_{11}$	GGACTAACAGTTCTACTGCCAT	TCTAAATAGTGCAGATTTGTGG	56	150–650	P	4–12	0.71
7	SEGMS36	$(GA)_{15}$	TAATATGCCAGTGTTGGTTAAA	GTGACATCTGACTGCTCCTAC	60	100–600	P	4–13	0.78
8	SEGMS47	$(TC)_{15}$	TGCTTAGATCTGGACTAAACCT	ATATCCTCTCATCTTCATCTGC	55	100–1,000	P	7–15	0.82
9	SEGMS55	$(CAA)_{10}$	TTTACAACATTAGAGTCCGTCA	TAAGGAAGGACAATTCTAAAGG	56	50–800	P	5–12	0.77
10	SEGMS71	$(TGT)_{12}$	GAATTCCTAGTGATTCTCTTGC	GGTACATCACTCATCTAGGACC	58	100–1,000	P	3–6	0.67
11	SEGMS122	$(AC)_{12}$	CATGATGAACAAAGTAAATTGC	AAGGTAAAGTGAGTACAAGGGA	61	200–800	P	6–14	0.80
12	SEGMS147	$(ACA)_{16}$	CATATGCTATCTGCAAATGTCT	CTTCTCTCTCTCTTATCCCGTA	58	100–700	P	4–13	0.81
13	SEGMS151	$(TGC)_{13}$	GACTCTCTTGCTTAGACTCTGG	GTTAATGGTTTAAGTGTGGTGG	55	200–550	P	5–14	0.80
14	SEGMS152	$(ATGC)_{5}$	GTATGATCTAGTCGTCAGGCTC	TGGCTCATCTATATTCACAATG	55	150–700	P	6–15	0.82
15	SEGMS155	$(TATC)_{6}$	TGAATTTACAGTCCATGAGGTA	ATCACGATTCTCTTGCTTAGAT	55	100–1,000	P	5–13	0.78
16	SEGMS163	$(GT)_{10}$	GATTTTCATGGTATAACGAGTCT	TCTCTAATACAATGAAACGCAG	56	100–1,000	P	6–14	0.80
17	SEGMS177	$(GTA)_{7}$	GTCGTTGTAAATGTCGTTGTAG	TTCTGCTTGAGTTAGTGAACTG	58	50–800	P	4–6	0.70
18	SEGMS179	$(CAA)_{7}$	GGTGAGTACATGGAGGTAAATC	CGAGAGGGTTCAGACTTAGTAG	55	50–700	P	4–7	0.70
19	SEGMS214	$(GTT)_{12}$	GAAGGCCCCCAACAATGTTGAT	TAACTAATTAAGCAAGCCACCT	55	150–1,000	P	3–6	0.68
20	SEGMS240	$(AAC)_{12}$	CAACAAGAAAGAACACAATGAG	AATGTGAATAACTAGGTGGTCC	55	100–900	P	7–14	0.83
21	SEGMS249	$(GAA)_{11}$	CATGAAGAAGAAAGGATATGGT	CTCATCATCACTAGAGCTTTCA	56	150–750	P	4–7	0.72
22	SEGMS355	$(AAG)_{7}$	ACCGGTACATATAGACAACCAG	AGGATGAATCCTATAGGTGATG	60	50–800	P	8–15	0.77
23	SEGMS726	$(AG)_{12}$	GAAGGAGAGAGGGAGTAGAGAG	TATAACATCGACACATCACCTC	56	100–950	P	4–8	0.74
24	SEGMS840	$(CT)_{10}$	TCGGATTCTCTGATTGTATGTA	ACTATTGATGAACTAACCGTCC	56	200–700	P	4–14	0.78
25	SEGMS926	$(AGA)_{9}$	TGATAGTGATGATGAAGATGCT	TGATTAGTTCATCGGTAGTGAA	58	100–850	P	5–12	0.77
26	SEGMS972	$(GTT)_{15}$	GCGCGACCGAACGATCACCA	CTATCTAACGAACGAACAAACA	61	100–700	P	6–13	0.80
27	SEGMS1035	$(CTT)_{25}$	ACTAGTACGTAGCTTCACATGAT	GAGAAAGGAAGGGAAGAAAG	55	50–450	P	4–12	0.81
28	SEGMS1069	$(AGA)_{10}$	CGGGAATTCGATTTCATGGGTTC	GACTACTCTAAGCATCGTCCTC	55	100–600	P	7–15	0.82
29	SEGMS1047	$(AT)_{8}$	ACTAGTGGACACGTAGACATGA	GGACAAGTGTAGATTGGAAGAT	55	100–750	P	5–12	0.78
30	SEGMS112	$(GTC)_{4}$	TTGTTGTTGTAGTAGTCGTCGT	CTAGTTCTTCCTACCTGCTATC	57	100–800	M	3–5	0.0
31	SEGMS116	$(GT)_{7}$	CGATTACTGATGTTCTTGATAACT	TAGTGATTCACACACACACACA	56	50–700	P	4–8	0.68
32	SEGMS580	$(GTA)_{4}$	GACTGTAACGAGCGGTAGAGAT	GCAGTGACTTCTACAACTACGA	56	50–450	P	4–8	0.69
33	SEGMS201	$(TGA)_{5}$	ACAAGAAAGAACAGGAGAAGAA	CTTGATATCCACTAGCATCAAA	55	50–500	P	7–14	0.79
34	SEGMS218	$(TTC)_{5}$	CATATCTTCACACTTGAAGCAC	GTTGAAGAAAGAAGAAGAGAGG	55	100–1,000	P	7–13	0.78

Table 1 continued

Sl. no.	SEGMS markers	Repeat motifs	Forward primer sequences (5′–3′)	Reverse primer sequences (5′–3′)	Annealing temperature (T°C)	Size (bp) range of fragments amplified	Polymorphic (P)/ monomorphic (M)	Range of fragments amplified	PIC value
35	SEGMS210	$(CAA)_5$	TTTCATGACAAAGCAATAATGT	GAGAAGAGGTAGAGGAAGACCT	56	150–1,000	P	6–14	0.76
36	SEGMS254	$(TCA)_5$	GAATCCAGCTAGTCATACTGGT	ATTACTGTGGTCCTTCGTAACT	58	100–550	M	3–7	0.0
37	SEGMS256	$(GC)_7$	ACACCGATCGCTCGTTCTCG	AGTCTGTCTGTTGCGTGTTT	61	50–950	P	7–12	0.77
38	SEGMS375	$(GAC)_4$	TAGATCTGGACTAATGAGGCTT	TTACTTGTCCTACAATCATCCA	58	100–600	P	3–6	0.68
39	SEGMS390	$(TGA)_5$	GTTGCTCTGAAACTATAGAGGG	CACCATATCCTTTCTTCTTCAT	55	100–500	M	3–5	0.0
40	SEGMS400	$(AGA)_5$	CACCTAGTTATGGGAACTTCAT	GAGTACCTTTACTGTCGTCCTC	55	50–900	P	5–8	0.74
41	SEGMS786	$(AAC)_5$	CTCTTCTTTCTTGTTCCTTCAC	TATTCTTAGTGGAAATCTTGGG	55	50–1,000	P	6–8	0.76
42	SEGMS940	$(AG)_6$	AGAGAGAGAGCATCGACTAAAG	TTGTACTATTACTGTTAGCGGC		100–800	P	3–6	0.69
43	SEGMS989	$(CT)_6$	GAGCTGCACTCTCTCTCTACTC	TGCAACGACACTAGTACTGAC	56	150–1,000	P	5–10	0.77
44	SEGMS1029	$(CCT)_5$	TAACATCATCACCAATCATTTC	GACGTCGGGCCCAACTGGACGCG	56	50–900	P	6–9	0.76

from the *Saccharum* species and varieties, were grouped in a separate cluster (III). The relationship among the varieties based on the two microsatellite types however, did not show complete correspondence.

Discussion

The pre-cloning selective hybridization based microsatellite enrichment technique is a robust, reproducible and cost-effective approach for isolating large number of microsatellites from diverse plant species with higher efficiency. The level of enrichment of sugarcane genomic libraries for different microsatellite motifs in the present study was 22%. This enrichment rate was higher than the range (10–20%) observed for similar enriched libraries in many crop plants, particularly in the graminaceous species, namely sugarcane (Cordeiro et al. 1999), rice (Chen et al. 1997), wheat (Pestsova et al. 2000), maize (Sharopova et al. 2002), and *Sorghum* (Bhattramakki et al. 2000).

Trinucleotide repeat-motifs were the most prevalent class of microsatellite among the SEGMS identified in sugarcane primarily because multiple probes with trinucleotide repeat motifs were used during the enrichment process. Higher frequency of $(CA)_n$ dinucleotide repeat-motifs than $(GA)_n$ in the SEGMS is comparable to those observed in earlier studies of sugarcane (Cordeiro et al. 2000), but not in line with other grass species (Powell et al. 1996; Chen et al. 1997), where the $(GA)_n$ rich motifs were reported to be abundant. This reflected a significant deviation of sugarcane from its other family members. The microsatellites with longer repeat-motifs that possibly result from high length dependent replication slippage have a higher tendency to be polymorphic (Temnykh et al. 2001). The proportion of perfect class I repeat-motifs was found in 35.7% of the total SEGMS identified, which suggested that the markers designed in this study from the sugarcane genomic libraries included a significant proportion of class I repeat-motifs and thus would have greater utility in sugarcane. Interestingly, 23.1% of the microsatellite containing sugarcane genomic sequences had significant sequence homology to retrotransposons. This compares with earlier observations on the ubiquitous and abundant distribution of *Copia* and *Gypsy* types LTR retrotransposons in sugarcane (36%, Rossi et al. 2001), maize (32%, Bennetzen 2000), and rice (13%, IRGSP 2005) genomes. The insertion and accumulation of retrotransposons between the genes possibly have played a major role in plant genome expansion (Bennetzen 2000). Thus, their wide genomic distribution and high degree of sequence variation in the large complex sugarcane genome (Rossi et al. 2001) are expected. This suggested the potential use of retrotransposon associated SEGMS repeat-motif

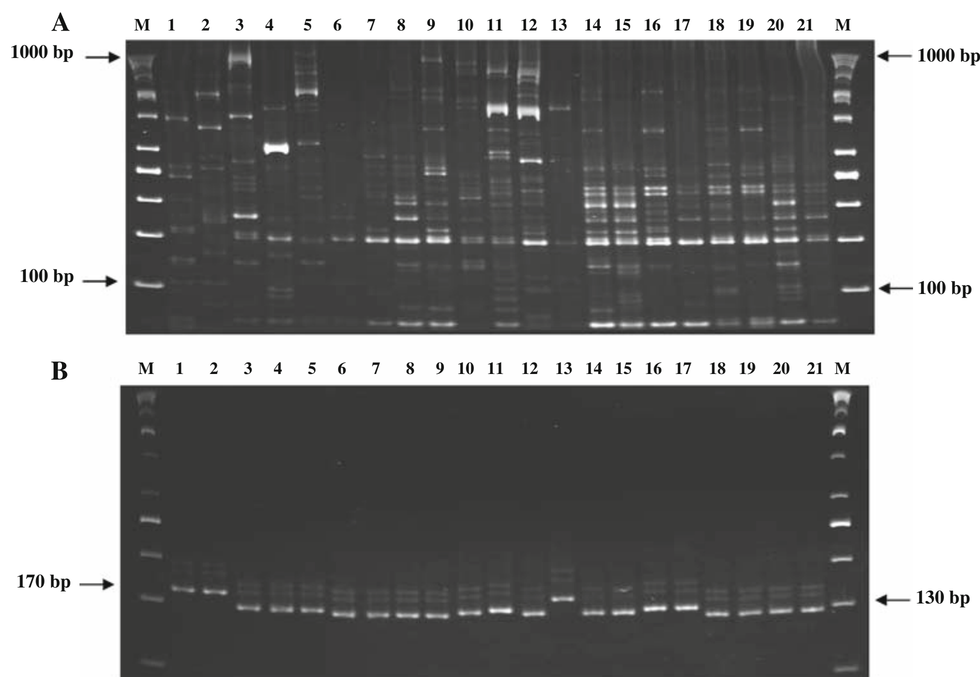

Fig. 2 Amplification profiles of class I SEGMS (**a**) and genic (**b**) microsatellite markers containing $(TGT)_{11}$ and $(TA)_{21}$ repeat-motifs, respectively, in 21 genotypes belonging to five cereal species and five sugarcane species clones, three related genera and eight commercial varieties. *Lanes 1* barley, *2* wheat, *3* rice, *4* maize, *5 Sorghum*, *6 S. officinarum*, *7 S. barberi*, *8 S. sinense*, *9 S. robustum*, *10 S. spontaneum*, *11 Narenga*, *12 Sclerostachya*, *13 Erianthus*, *14–21* commercial sugarcane varieties Co 419, Co 8021, Co 8371, CoPant 84212, CoS 8436, Co 1148, Co 62175 and Co 86249. The banding pattern obtained with the SEGMS markers was far more complex than that generated with the UGMS markers. Amplicons were resolved in 10% native PAGE. *M* 50 bp DNA ladder. *Arrows* indicate the polymorphic amplicons

sequences as a source to develop polymorphic class of genetic markers in sugarcane.

Primer-pairs could be designed from the flanking genomic sequences for 90% of the identified microsatellite motifs in sugarcane. Our failure to design primers for the rest (10%) was due to different attrition factors such as failure of sequencing past the microsatellite motifs (4%, 51 out of 1,261 clones) and/or location of repeat-motif tract too near to the clone insertion sites (5.1%, 65), thus offering little or no flanking regions for designing primers. The clone exclusion and hindrances observed during the construction of sugarcane enriched genomic library have been commonly observed earlier in other crop species (Ramsay et al. 2000; Bhattramakki et al. 2000). Earlier, Cordeiro et al. (2000) reported that only 27% of the total microsatellite containing genomic clones identified from the sugarcane enriched library had primer designing potential. In several other studies the primer designing potential of enriched genomic clone containing microsatellites varied from 51% in wheat (Pestsova et al. 2000) to 66% in *Sorghum* (Bhattramakki et al. 2000). Higher efficiency in the present study could be due to appropriate insert size selection for enrichment and generation of longer high-quality sequences through bidirectional sequencing of microsatellite enriched genomic clones, thus offering enough flanking sequences to design primers for amplification of individual microsatellite loci.

The utility of SEGMS markers was evaluated by their potential to amplify the target sequence and detect polymorphism. The primers designed from the sequences flanking the microsatellite repeat-motifs were found highly efficient in giving amplification with a success rate of 87.4%. This was significantly higher than the amplification efficiency reported earlier for similar enriched genomic libraries in wheat (49%, Roder et al. 1998), but marginally higher/comparable to sugarcane (80%, Cordeiro et al. 2000) and *Sorghum* (85%, Bhattramakki et al. 2000). The genomic sequences from which primers were designed had balanced (52–53%) GC content, which possibly supported greater primer binding (Temnykh et al. 2001), and thus provided higher amplification efficiency of SEGMS markers in sugarcane.

Further, 93.2% (PIC of 0.75) of the markers designed in this study had potential of showing polymorphism among the *Saccharum* species, related genera, and varieties. This is higher than the level reported earlier with the microsatellite markers derived from the enriched genomic libraries of sugarcane (91%, Cordeiro et al. 2000; 35%, Pinto et al. 2004; 72%, Aitken et al. 2005), maize microsatellites (46%, Selvi et al. 2003), and random markers

Table 2 Comparison of the cross-transferability and polymorphic potential of the SEGMS markers with that of UGMS markers

Microsatellite markers	Source of sequence	Number of markers used			Number of fragments amplified			Polymorphism observed			Mean PIC	Polymorphic potential						Cross-transferability to five cereals and *Saccharum* complex Total[a]
												Inter-generic		Inter-specific		Inter-varietal		
		Class I	Class II	Total	Minimum	Maximum	Average	Class I[a]	Class II[a]	Total[a]		Per cent polymorphism[a]	PIC	Per cent polymorphism[a]	PIC	Per cent polymorphism[a]	PIC	
SEGMS	Genomic sequences	28	16	44	3	15	9	28 (100)	13 (81.2)	41 (93.2)	0.75	39 (88.6)	0.76	36 (81.8)	0.73	33 (75)	0.79	26 (59)
UGMS	Unigene sequences	21	–	21	1	4	2.5	8 (38)	–	8 (38)	0.52	7 (33.3)	0.57	7 (33.3)	0.57	4 (19)	0.52	20 (95.2)

[a] The number in the bracket is the percentage of markers showing polymorphism

such as RAPD (67%, Nair et al. 2002) and AFLP (71%, Selvi et al. 2006). The extent of polymorphism as detected by SEGMS markers was comparatively higher than the microsatellite markers derived from the unigene sequences (38%). It could be due to the use of highly conserved expressed coding sequences as against less conserved genomic sequences, which are under moderate selection pressure (Cho et al. 2000). We observed that the SEGMS markers, particularly for the longer class I tetranucleotide and dinucleotide repeat-motifs showed higher potential (100%) to detect polymorphism as compared to those from the class I trinucleotide (90.3%) and class II motifs (81.2%). This reflected a correlation between the type and length of repeat-motifs with the level of polymorphism as reported earlier in rice (Temnykh et al. 2001) and thus provided a strategy for rapid selection of more polymorphic markers for genotyping applications in sugarcane. Sequencing of SEGMS fragment length variants at three microsatellite loci revealed that the presence of varied copy number of repeat-motifs in different amplicons was the primary basis of SEGMS fragment length polymorphism. Such fragment length polymorphism at microsatellite loci due to large variation in the copy number of repeats have been reported earlier in rice (Cho et al. 2000), maize (Matsuoka et al. 2002), and chickpea (Sethy et al. 2006).

The extent of cross-transferability of a marker system determines its suitability in comparative genome mapping and phylogenetics. Striking differences in the levels of cross-transferability observed for SEGMS (25%) and UGMS (95%) markers to cereals was most likely because of assay of different regions of the genome by these two markers (Cho et al. 2000; Pinto et al. 2006). UGMS markers being derived from coding region of the genome are expected to be more conserved and thus more cross-transferable. However, high (93.2%) cross-transferability of SEGMS markers to *Saccharum* species and related genera suggested the utility of these markers in mapping of genes from related species and genera, identification of true inter-generic and inter-specific hybrids, monitoring of introgression from the unadapted relatives to the cultivated genetic backgrounds, comparative mapping, and establishing evolutionary relationship.

The SEGMS markers detected a wider range of genetic diversity (0.16–0.82; an average of 0.44) among the *Saccharum* species, related genera and Indian commercial sugarcane varieties than the unigene derived microsatellites used in this study (0.50–0.80; 0.67) as well as the other markers, namely RAPD (0.59–0.81; 0.71; Nair et al. 2002), maize microsatellites (0.40–0.73; 0.64; Selvi et al. 2003), and AFLP (0.52–0.83; 0.62; Selvi et al. 2005), which were employed earlier using a larger set of species clones and Indian varieties. It thus revealed higher efficiency of the SEGMS markers and suggested that a smaller set of

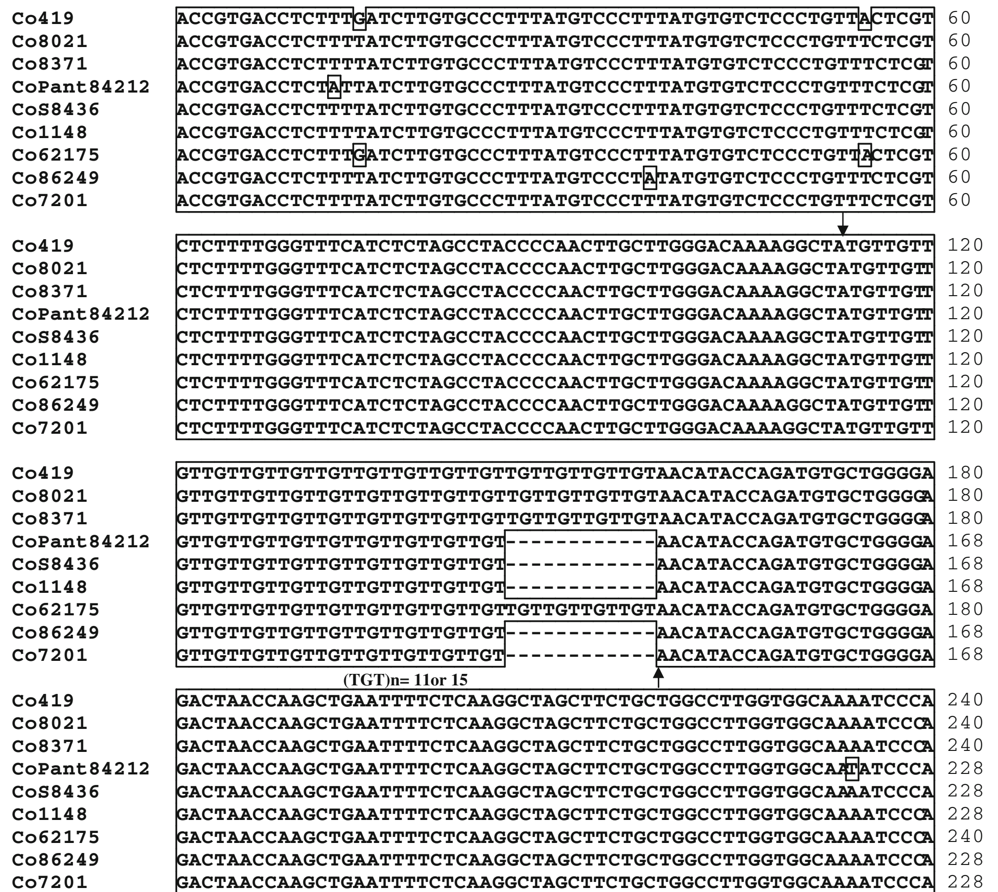

Fig. 3 Multiple sequence alignment of the size variant fragments amplified from Co 7201 and eight other sugarcane varieties for the SEGMS33 marker showing the presence of microsatellite repeat-motif, $(TGT)_n$ with 11 repeat-units in CoPant 84212, CoS8436, Co 1148 and Co 86249 and 15 repeat units in Co 419, Co 8021, Co 8371 and Co 62175. The repeat motifs are conserved; however, the presence of SNPs in the flanking genomic sequences is highlighted. *Arrows* indicate the repeat-motif sequences

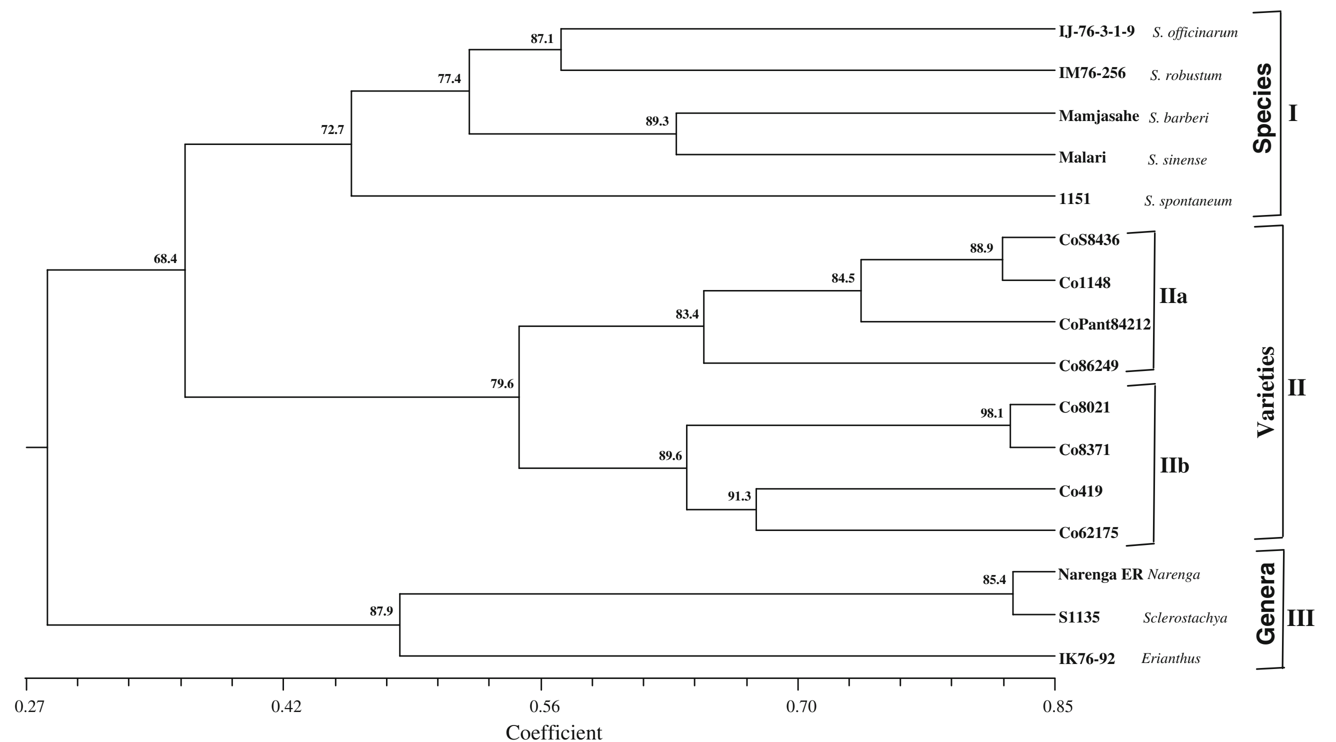

Fig. 4 Dendrogram showing genetic relationship among the sugarcane species clones, related genera and Indian sugarcane varieties based on Jaccard's similarity coefficient using 44 SEGMS markers. The bootstrap values are indicated at the *nodes* in each cluster

SEGMS markers can be effectively used for assaying more relevant regions of large and polyploid sugarcane genome for revealing higher level of genetic diversity. Correspondence of genetic relationship established by these markers with the expectations based on taxonomic and pedigree relationship (Daniels et al. 1975; D'Hont et al. 1993) suggested that the genetic diversity assayed by the SEGMS markers was realistic and thus would be of use in sugarcane breeding when applied to a larger set of germplasm lines.

Genetic analysis has been hindered in sugarcane because of its complex polyploid genome and lack of sufficiently informative markers. A large number of useful polymorphic markers are thus required for mapping the whole genome, finding markers tightly linked to target genes and monitoring introgression for broadening the genetic base of sugarcane varieties. Our results suggested that the SEGMS markers developed in this study would be of immense use for various genotyping applications in sugarcane.

Acknowledgments The work presented in the manuscript was funded by the Department of Biotechnology (DBT), Government of India. The authors thank the Institute of Plant Genetics and Crop Research (IPK) for the availability of microsatellite search tool MISA.

References

Aitken KS, Jackson PA, McIntyre CL (2005) A combination of AFLP and SSR markers provides extensive map coverage and identification of hom(oe)ologous linkage groups in sugarcane. Theor Appl Genet 110:789–801

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Bennetzen JL (2000) Transposable element contributions to plant genome evolution. Plant Mol Biol 42:251–269

Bhattramakki D, Dong J, Chhabra K, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43:988–1002

Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor Appl Genet 95:553–567

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, Park WD, Ayres N, Cartinhour S, McCouch SR (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Cordeiro GM, Maguire TL, Edwards KJ, Henry RJ (1999) Optimization of a microsatellite enrichment technique in *Saccharum* spp. Plant Mol Biol Rep 17:225–229

Cordeiro GM, Taylor GO, Henry RJ (2000) Characterization of microsatellite markers from sugarcane (*Saccharum* spp.), a highly polyploid species. Plant Sci 155:161–168

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross-transferable to *Erianthus* and *Sorghum*. Plant Sci 160:1115–1123

D'Hont A, Paulet F, Glaszmann JC (1993) Oligoclonal interspecific origin of 'North Indian' and 'Chinese' sugarcanes. Chromosome Res 10:253–262

Daniels J, Smith P, Panton N, Williams CA (1975) The origin of the genus *Saccharum*. Sugarcane Breed Newsl 36:24–39

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758–760

Goldstein DB, Schlotterer C (1999) Microsatellites: evolution and applications. Oxford University Press, Oxford

IRGSP (2005) The map based sequence of rice genome. Nature 436:793–800

Jaccard P (1908) Nouvelles recherché sur la distribution florale. Bull Soc Vaudoise Sci Nat 44:223–270

Liu ZW, Biyashev RB, Saghai-Maroof MA (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 93:869–876

Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J (2002) Micosatellites in *Zea*-variability, patterns of mutations, and uses for evolutionary studies. Theor Appl Genet 104:436–450

Nair NV, Selvi A, Srinivasan TV, Pushpalatha KN (2002) Molecular diversity in Indian sugarcane cultivars as revealed by randomly amplified DNA polymorphisms. Euphytica 127:219–225

Parida SK, Rajkumar KA, Dalal V, Singh NK, Mohapatra T (2006) Unigene derived microsatellite markers for the cereal genomes. Theor Appl Genet 112:808–817

Pestsova E, Ganal MW, Roder MS (2000) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43:689–697

Pinto LR, Oliveira KM, Ulian EC, Garcia AA, deSouza AP (2004) Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome 47:795–804

Pinto LR, Oliveira KM, Marconi T, Garcia AAF, Ulian EC, deSouza AP (2006) Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed 125:378–384

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Ramsay L, Macaulaya M, degli Ivanissevichb S, MacLeana K, Cardlea L, Fullera J, Edwardsc KJ, Tuvessond S, Morganteb M, Massarie A, Maestrie E, Marmirolie N, Sjakstef T, Ganalg M, Powella W, Waugha R (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997–2005

Roder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Rossi M, Araujo PG, Sluys MV (2001) Survey of transposable elements in sugarcane expressed sequence tags (ESTs). Gen Mol Biol 24:147–154

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

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

Selvi A, Nair NV, Noyer JL, Singh NK, Balasundaram N, Bansal KC, Koundal KR, Mohapatra T (2005) Genomic constitution and genetic relationship among the tropical and subtropical Indian sugarcane cultivars revealed by AFLP. Crop Sci 45:1750–1757

Selvi A, Nair NV, Noyer JL, Singh NK, Balasundaram N, Bansal KC, Koundal KR, Mohapatra T (2006) AFLP analysis of the phenetic organization and genetic diversity in the sugarcane complex, *Saccharum* and *Erianthus*. Genet Resour Crop Evol 53:831–842

Sethy NK, Shokeen B, Edwards KJ, Bhatia S (2006) Development of microsatellite markers and analysis of intraspecific genetic variability in chickpea (*Cicer arietinum* L.). Theor Appl Genet 112:1416–1428

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Davis G, Coe EH (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481

Temnykh S, Declerk G, Lukashover A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length-variation, transposon associations and genetic marker potential. Genome Res 11:1441–1452

Yap I, Nelson RJ (1996) Winboot: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. IRRI Discussion paper series no. 14. International Rice Research Institute, Manila, Philippines