

Development of microsatellite markers targeting (GATA)_n motifs in sorghum [*Sorghum bicolor* (L.) Moench]

I. Jaikishan · G. R. Paik · R. Madhusudhana ·
M. Elangovan · P. Rajendrakumar

Received: 12 October 2011 / Accepted: 16 July 2012 / Published online: 28 July 2012
© Springer Science+Business Media B.V. 2012

Abstract Microsatellite markers targeting (GATA)_n motifs are known to be highly polymorphic. Genome-wide development of such markers has not been reported in sorghum. The main objective of this study was to identify Class I microsatellites with (GATA)_n motifs in the sorghum genome through in silico analysis and assess their potential as molecular markers. The study identified a total of 128 such motifs, of which 14, 16 and 98 motifs were present in the genic, upstream and non-genic regions, respectively. The majority of the (GATA)_n motifs were found in the non-genic regions of the genome while 23.44 % of them were found within the genes and upstream of genes. About 110 PCR-based markers were developed targeting these microsatellites and 50 of them distributed across the genome were validated in 24 diverse sorghum genotypes representing different racial groups. Thirty-eight markers were polymorphic, with average polymorphism information content value of 0.69, and the sorghum genotypes could be grouped into two major clusters. These markers with robust amplification combined with good allelic diversity represent a new set of microsatellite markers in

sorghum reported for the first time that will be highly useful for various genetics and molecular breeding applications.

Keywords Microsatellites · Sorghum · (GATA)_n motifs · PCR-based markers · Genetic diversity

Sorghum [*Sorghum bicolor* (L.) Moench] is an important staple food and fodder crop in tropical/semi-tropical African and Asian regions (Doggett 1998). All cultivated sorghums belong to *S. bicolor* ssp. *bicolor*, which is further divided into five races (*bicolor*, *caudatum*, *guinea*, *durra*, *kafir*) based on morphological characters. There are also 10 intermediate races resulting from all possible inter-racial crosses (Harlan and de Wet 1972). Information on genetic diversity in the available germplasm is essential in deciding their utility in crop breeding programmes. Earlier, morphological characteristics (Dje et al. 1998) and isozymes (Danquah et al. 2000) were employed for the assessment of genetic diversity in sorghum. These approaches require more time and expertise to efficiently classify material and are based on limited traits. Moreover, the results are subjective, due to environmental influences. This situation demands the development of robust molecular marker technology which will help in a better estimation of genetic diversity.

Among the molecular markers, microsatellites are the most favored for a variety of applications in crop genetics and breeding because of their multi-allelic

Electronic supplementary material The online version of this article (doi:10.1007/s11032-012-9770-3) contains supplementary material, which is available to authorized users.

I. Jaikishan · G. R. Paik · R. Madhusudhana ·
M. Elangovan · P. Rajendrakumar (✉)
Directorate of Sorghum Research, Hyderabad 500030,
Andhra Pradesh, India
e-mail: rajendra@sorghum.res.in

nature, high reproducibility, co-dominant inheritance, abundance and extensive genome coverage (Gupta and Varshney 2000). The usefulness of microsatellite markers in genetic diversity analysis (Shehzad et al. 2009; Wang et al. 2009) and quantitative trait locus (QTL) mapping (Parh et al. 2008; Wu and Huang 2008) has been amply demonstrated in sorghum. The availability of the complete sequence of the sorghum genome in the public domain offers the opportunity for the identification of motifs of interest through in silico approaches, thereby helping in the rapid development of PCR-based markers. Earlier studies have revealed that (GATA)_n motifs are abundant in many eukaryotic genomes (Singh 1995) and several such motifs serve as transcription factors (Trembley and Viger 2003). Even though the (GATA)_n motif-based microsatellite markers are widely used for paternity testing, genetic counseling and identification of individuals in humans (Pena et al. 1994), their utility in plant species is very limited. A small number of reports in crops such as pearl millet (Chowdari et al. 1998), rice (Davierwala et al. 2001), papaya (Parasnis et al. 1999), chickpea (Serret et al. 1997), sunflower (Mosges and Friedtu 1994) and tomato (Grandillo and Tanksley 1996; Rao et al. 2006) have highlighted the potential of (GATA)_n motifs in analyzing the genetic variation. Recently, microsatellite markers targeting (GATA)_n motifs across the genome were developed in rice and their utility in genetic diversity assessment was demonstrated (Rajendrakumar et al. 2009). Such markers have not been reported in sorghum. The objectives of this study were to analyze the abundance and distribution pattern of Class I (GATA)_n motifs in sorghum, develop PCR-based markers and demonstrate their utility in genetic diversity analysis.

The whole genome sequence of sorghum (*S. bicolor* cv. Btx623) was downloaded from the Joint Genome Institute database (Paterson et al. 2009; ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Sbicolor/). Class I microsatellites (motifs with repeat length ≥ 20 nt; McCouch et al. 2001) with (GATA)_n motifs were identified in the sorghum genome using a Perl script developed in-house. The identified (GATA)_n motifs were designated as genic (exons and introns), upstream (within 2 kb from transcription start site) and non-genic based on the results of BLASTN analysis of the motif as query sequence using the gene annotation information available at the Phytozome database (<http://www.phytozome.net/sorghum>). Primers were

designed targeting all the Class I (GATA)_n motifs (Supplementary Table 1) using the web resource Primer3 (<http://frodo.wi.mit.edu/primer3/>). Primer pairs targeting 50 motifs (five per chromosome) with maximum repeat lengths were used for further analysis and were physically mapped in the sorghum genome based on their positions using the software MapChart 2.2 (Voorrips 2002). Through a preliminary analysis, 48 primer pairs which showed good amplification were selected and then used for amplification of DNA isolated from 24 sorghum genotypes (Supplementary Table 2) representing different races. Total genomic DNA was extracted using the CTAB method (Saghai-Marouf et al. 1984) and used for PCR amplification according to the procedure described by Panaud et al. (1996) with modifications. PCR was carried out using 20 ng DNA as template for amplification, 2 pmol of each primer, 0.1 mM dNTPs, 1 \times PCR buffer (Bangalore Genei, India), and 1 unit of *Taq* DNA polymerase (Bangalore Genei, India) in a total volume of 10 μ l. The PCR cycling conditions followed were: a 5 min initial denaturation at 94 °C followed by 35 cycles of 30 s denaturation at 94 °C, 30 s annealing at 55 °C, 1 min extension at 72 °C, concluding with a final extension of 72 °C for 7 min. PCR products were resolved on a Bio-Rad Sequi-Gen™ sequencing electrophoresis apparatus (Bio-Rad Laboratories, USA) in 6 % non-denaturing polyacrylamide gel. The DNA fragments were visualized by silver staining (Fritz et al. 1995). Amplicons were scored individually as ‘1’ and ‘0’ for the presence or absence, respectively, of an allele, and a binary data matrix was generated. The polymorphism information content (PIC) value for each microsatellite marker was calculated using the online resource ‘Polymorphism Information Content Calculator’ available at www.liv.ac.uk/~kempsj/pic.html. The binary data matrix was subjected to cluster analysis. Genetic diversity estimate-related analyses were done using NTSYSpc ver.2.02i (Rohlf 2000). Genetic similarities (GS) between pairs of accessions were measured by the DICE similarity coefficient based on the proportion of shared alleles (Dice 1945; Nei and Li 1979) with SIMQUAL module. The clustering of accessions was done based on a similarity matrix using an unweighted pair group method with arithmetic average (UPGMA) algorithm following SAHN module. The similarity result was then used to construct a dendrogram following TREE module. The reliability of the tree was tested by bootstrap

analysis using the software Winboot (Yap and Nelson 1996).

The present study identified a total of 128 Class I microsatellites with (GATA)_n motifs in the sorghum genome (Supplementary Table 3). Of these, 14 were in the genic region, 16 were present upstream of genes and the remaining 98 repeats were non-genic. Among the genic motifs, two were present in the 3' untranslated region (UTR), while 12 were present in introns. Only 23.44 % (30 out of 128) of the (GATA)_n motifs in sorghum may have functional significance due to their presence in genic and upstream regions, which is lower than of the proportion in rice (51.03 %) reported by Rajendrakumar et al. 2009. The presence of a lower number (128) of Class I (GATA)_n motifs in the sorghum genome than in the rice genome (243) may be attributed to the presence of gaps in the sorghum genome sequence and also due to the more recent evolution of rice as compared to sorghum. The details of the distribution of these motifs are given in Supplementary Table 4. The frequency of (GATA)_n motifs was high in relatively smaller chromosomes (Chromosome 5 and 7) while the frequency was low in a larger chromosome (Chromosome 2). This observation clearly indicated the lack of correspondence between the number of (GATA)_n motifs and the chromosome size, which was similar to that observed in human and primates by Subramanian et al. (2003). It is interesting to note that the (GATA)_n motifs show a skewed distribution and are mostly found in clusters, with a majority of them (37.50 %) towards the telomeric region of the chromosome; chromosome 9 had the least motifs (12.5 %) while chromosome 6 and 7 had the most motifs (55.55 %) in the telomeric region. This may be due to the conservation of sequences in the telomeric regions. This observation is in contrast to that of rice where these motifs were found as clusters distributed across the length of chromosomes (Rajendrakumar et al. 2009) and in tomato where these motifs were clustered around the centromere (Areshchenkova and Ganai 1999; Grandillo and Tanksley 1996).

About 14 Class I (GATA)_n motifs were identified in the genic region of the sorghum genome, of which two were present in the 3' UTR and 12 were intronic. One of the motifs in the 3' UTR was present in the gene encoding monosaccharide transport protein 1 while the other motif was present in the gene encoding expressed protein (Supplementary Table 4). The

presence of microsatellite motifs in genic regions could have implications in the molecular function of the encoded protein. The presence of (GATA)_n motifs in genic regions of metazoan eukaryotic genomes and its implications in post-transcriptional signaling of mRNAs encoding membrane-associated proteins was reported by Riley and Krieger (2002). Interestingly, a clear association of (GATA)_n motifs with the sex of papaya was reported by Parasnis et al. (1999). The presence of these motifs in the genes encoding for disease resistance in rice was reported by Davierwala et al. (2001).

Twelve (GATA)_n motifs were present in the intronic region of genes. Nine of these motifs were present in genes encoding proteins with known functions while three motifs were present in genes with unknown function (Supplementary Table 4). The genes coding for proteins with known function include putative CLB1 protein (calcium-dependent lipid binding) protein, major facilitator superfamily protein, putative peroxidase 1, putative serine-type endopeptidase inhibitor, putative transcription activator RF2a, SBP-domain protein 1, auxin efflux carrier component 1, caffeic acid 3-*O*-methyltransferase and cytochrome P450 71E1. Caffeic acid *O*-methyltransferase gene is responsible for brown mid-rib trait (*bmr12*) in sorghum and was cloned through a candidate gene approach by Bout and Vermerris (2003). The presence of (GATA)_n motifs in the intronic regions of the gene may be associated with differential splicing of mRNA and determination of splice junctions similar to that reported in rice (Rajendrakumar et al. 2009). In humans, a GATA motif present in the conserved region within α -synuclein gene (*SNCA*) intron-1 associated with Parkinson's disease directly induces a 6.9-fold increase, thereby involving in gene regulation (Scherzer et al. 2008).

A total of 16 (GATA)_n motifs were identified in the upstream regions of genes and six of these code for putative uncharacterized proteins while the remaining 10 code for proteins with known as well as unknown functions. GATA sequences present upstream of genes have been associated with gene regulation (McNeil et al. 2008; Trembley and Viger 2003). The majority of the Class I (GATA)_n motifs identified (111 motifs) were present in non-genic regions (Supplementary Table 4).

PCR-based markers were developed by designing primer pairs targeting all the Class I (GATA)_n motifs

Table 1 Details of polymorphic (GATA)_n motif-based microsatellite markers used in the study

Marker	Motif	Primer F	Primer R	Chromosome no.	Physical position (Mb)	Location of motif	N _a	PIC
SbGM1-7	(GATA) ₁₁	ATGAATATTGGGGTGAGAGATG	TCTATCTATCTACTCTCCGTCCTG	1	16.7	Non-genic	8	0.83
SbGM1-16	(GATA) ₁₈	TAAAGCTTTC AAGGACATAGGG	ATGTGATTCCTGTTCCTGTGG	1	69.4	Non-genic	7	0.81
SbGM1-20	(GATA) ₁₅	GTGTTGAGCTGTGAGTGAGC	TCCCTACTAGAACTAGAACTCC	1	73.4	Non-genic	6	0.71
SbGM2-2	(GATA) ₈	TAAAGTAACTCATCTCAAGC	GTTGAACTTTGTAGCAAAACAGG	2	0.45	Intronic	3	0.50
SbGM2-6	(GATA) ₁₉	AAAGAAAGAAGGACTAGAGG	GATGCTTTGGTCTTTTCTCG	2	9.4	Non-genic	5	0.72
SbGM2-10	(GATA) ₂₉	TGAGTTCAGTGACATAAGAGACG	TGCTTTGAGTTTGAAGAGTTAGC	2	63.8	Non-genic	8	0.79
SbGM3-1	(GATA) ₁₈	GTTCAITTTGAAGGGTCATCC	ATAAAGTGGCTGGCTGTGG	3	0.03	Non-genic	6	0.73
SbGM3-2	(GATA) ₂₆	GGCAATCCTGTATGATAGAAC	ACACACGCACATCGTAGTTTC	3	10.1	Exonic	9	0.83
SbGM3-4	(GATA) ₆	AAACCGAGTTTAGTATCACAGC	GGGTATCGACTTCTGAATAAAGG	3	53.7	Upstream	3	0.26
SbGM3-7	(GATA) ₁₄	CTGGAGACGATGAAGATGC	AGGATCGTGTGTGTGTC	3	74.2	Non-genic	9	0.81
SbGM4-2	(GATA) ₁₈	ACATCTGATGACCCCTGACG	TATTCGATCATGTAGGGTTGC	4	5.04	Intronic	10	0.83
SbGM4-5	(GATA) ₁₅	ATCCTATGAAATCCTACAATGG	GGCTAAATATTGAGGTTGAAGG	4	54.1	Non-genic	7	0.79
SbGM4-9	(GATA) ₁₇	TAACTTCGCAACTGTCAAGC	ACTGAAATCAGTCCACACTCC	4	61.2	Non-genic	8	0.83
SbGM4-10	(GATA) ₈	CTCTCTCTCTGATCTTCTAGG	AGCTGGTTAGAACTTTTCTGG	4	66.4	Non-genic	4	0.58
SbGM5-1	(GATA) ₂₀	TTTCACATTC AAGGTCATGG	ACACAGGGATAACAGGATAAGG	5	2.6	Non-genic	8	0.80
SbGM5-3	(GATA) ₃₆	CCATGTCAITTTTAGTTTTTTCG	GATGAAAGTCTCTGTACTAAAGG	5	10.5	Non-genic	7	0.82
SbGM5-6	(GATA) ₃₇ (GATG) ₂ (GATA) ₃₁	ACGTATAGCTAGCCTGGTTTTG	ATGCTGTGTGTCTCCGTCCTAC	5	23.5	Non-genic	8	0.82
SbGM5-8	(GATA) ₉	ATAGATGGATGTCATGGATGG	GAGATCTGGGTTTATCTATCTCG	5	52.8	Non-genic	3	0.58
SbGM5-16	(GATA) ₃₃	ATCTTCAACCTTGATTCACACG	ATTAAATCGATCCTCTGTAGTGC	5	61.3	Non-genic	9	0.86
SbGM6-1	(GATA) ₆	GCTAGCTACCTCTACCAAGTCC	CTAGTTTCTTTTGCACCTTACC	6	1.4	Non-genic	5	0.49
SbGM6-5	(GATA) ₁₈	ACAAAAGCTTCGTAGCAAAAC	AGGCTTAGCCCTTTGCATTTCC	6	51.3	Non-genic	6	0.71
SbGM6-6	(GATA) ₈	AAAAACATGAACGAACACACG	ACAAAACGAAAGTGTACAGTGG	6	57.1	Non-genic	3	0.48
SbGM6-7	(GATA) ₈ CTA(GATA) ₇	GTTGTTCAITTTGTGCTTTTC	CACATAGTTAATCCACGTTGTCG	6	60.5	Non-genic	3	0.55
SbGM7-1	(GATA) ₁₆	GCTGACCATCATTTCAAATCC	AACCAACCGCTTTGATTAGG	7	1.4	Non-genic	9	0.84
SbGM7-4	(GATA) ₁₈	GACGAAACAGATGAATAGACG	TGATTGTTTGAAGTGTAGTGTGC	7	3.9	Non-genic	7	0.78
SbGM7-12	(GATA) ₇ GATG(GATA) ₂	TTTTTGGGACGAAAGGAGTA	GTTGACTTCTCTCGAAAACCGA	7	48.8	Non-genic	5	0.66
SbGM7-14	(GATA) ₁₉	AAACCATGCCTATATGGAACC	GCATTATATTACGGAACGAAAGG	7	56.1	Intronic	5	0.68
SbGM8-1	(GATA) ₁₄	GCTTACAGCTACTACACCATGC	CTGCTTACTGTACGTACTTGTGG	8	3.2	Exonic	9	0.86
SbGM8-2	(GATA) ₅	CACCGATAAATGTGTTCTACG	ATGGCTCTCTCAGATTACG	8	6.2	Intronic	3	0.41
SbGM8-6	(GATA) ₆	GATCCCAGGAAATGATTTCTTC	TTGGGTATATGTTGTCTCTTCC	8	53.7	Non-genic	6	0.54
SbGM9-1	(GATA) ₅	GAGAGGCTAITTTCAAATCC	GGGAACGGTAGAAGGAAAC	9	5.8	Non-genic	4	0.39

Table 1 continued

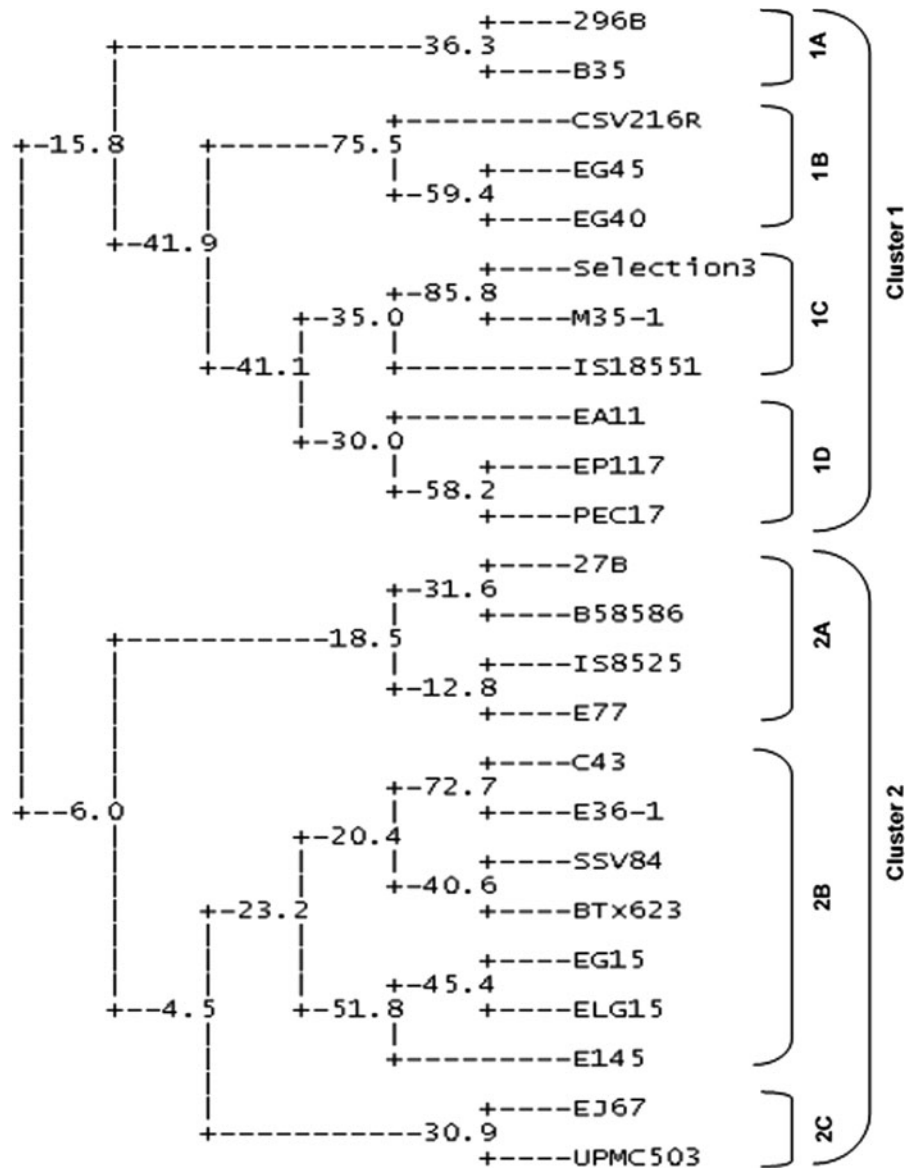
Marker	Motif	Primer F	Primer R	Chromosome no.	Physical position (Mb)	Location of motif	N_a	PIC
SbGM9-2	(GATA) ₇	GATATACCCGGTAGCTCTATTGC	AACCGTAGAGGATGTTTTCG	9	9.1	Non-genic	5	0.64
SbGM9-6	(GATA) ₁₂	TGAGAGTGATCGTGATTTTAGG	GGCCACAAGATAGAAAATGC	9	47.4	Non-genic	7	0.81
SbGM9-7	(GATA) ₁₃	GCTACAGAGGAATGTGTATGC	TCTCCTTGTCTATCTGACTATCTC	9	58.9	Non-genic	5	0.68
SbGM10-1	(GATA) ₁₃	TATAGAGAGACACCTGGGAAG	ATCTTATGTGCACGTCTACAAG	10	2.3	Non-genic	5	0.70
SbGM10-5	(GATA) ₁₂	GTAGAAAGAGAGGAGGGAGAGG	AAGCCCTAGC-AAATTTAGATCC	10	47.9	Non-genic	5	0.58
SbGM10-6	(GATA) ₁₆	TGTGTATATGATGCCAGATCC	GGGTATGCGATATAAAATGC	10	53.1	Non-genic	7	0.77
SbGM10-8	(GATA) ₁₃ ATA (GATA) ₄ TCTA(GATA) ₆	TGCCAGTGTCAATATAACTTCC	ATCACTGTGGATGAGACATCG	10	56.8	Non-genic	6	0.76
				Average			6.13	0.69

N_a Number of alleles, PIC Polymorphism information content

present in the sorghum genome. If two or more (GATA)_n motifs were found adjacent to one another, a single primer pair flanking these motifs was designed. Thus a total of 110 microsatellite markers were developed, which were distributed across 10 chromosomes (Supplementary Table 1). Primers could not be designed for eight motifs because of the presence of gaps in the flanking regions. The physical map location of the (GATA)_n motif-based microsatellite markers developed in this study is shown in Supplementary Fig. 1. To assess the potential of (GATA)_n motif-based microsatellite markers, 50 markers were selected as mentioned above and analyzed in a set of diverse sorghum genotypes. All the markers except SbGM1-5 and SbGM8-3 showed clear and robust amplification, of which 38 markers were polymorphic. A total of 233 alleles were generated by these 38 microsatellite markers among 24 genotypes. The number of alleles ranged from 3 to 10 with an average of 6.13 per marker. The marker SbGM4-2 amplified a maximum number of 10 alleles, while six markers (SbGM2-2, SbGM3-4, SbGM5-8, SbGM6-6, SbGM6-7 and SbGM8-2) amplified only three alleles among the genotypes studied. An average PIC value of 0.69 was recorded for these markers, with SbGM8-1 exhibiting a maximum PIC value of 0.86, while SbGM3-4 exhibited a minimum value of 0.26. The total number of alleles generated and the number of polymorphic alleles with their corresponding PIC values are listed in Table 1.

The average number of alleles per marker (6.13) detected in this study using 38 polymorphic microsatellite markers was higher than that detected by Schloss et al. (2002) on 25 sorghum lines (3.4), Ali et al. (2008) on 72 sorghum accessions (3.22), and Agrama and Tuinstra (2003) and Smith et al. (2000) with mean allele numbers per marker of 4.3 and 5.9, respectively. The gene diversity or PIC observed in this study (mean PIC value = 0.69) is higher than the gene diversity values (0.40, 0.46, 0.58, 0.62) reported by Ali et al. (2008), Schloss et al. (2002), Smith et al. (2000) and Agrama and Tuinstra (2003), respectively. It is interesting to note that (GATA)_n motif-specific markers exhibited good allelic diversity. This observation is similar to the high resolution power and good allelic diversity of (GATA)_n motif-based microsatellite markers reported in rice by Rajendrakumar et al. (2009). Most of these markers yielded sharp and robust amplification without multiple amplicons,

Fig. 1 Dendrogram constructed using $(GATA)_n$ motif-based microsatellite marker data revealing the clustering of sorghum genotypes



which may be due to the presence of limited numbers of $(GATA)_n$ motifs and their unique occurrence without multiple primer binding sites, similar to that reported in rice by Rajendrakumar et al. (2009) in rice. The $(GATA)_n$ motif-based microsatellite markers developed in this study can be considered as informative and unique and can supplement the microsatellite markers reported by many research groups and the genome-wide microsatellite markers reported by Yonemaru et al. (2009).

Based on the 233 polymorphic alleles, the Dice genetic similarity co-efficient was estimated for each

pair of the 24 sorghum genotypes, and ranged from 0.08 to 0.63 (Supplementary Table 5). The wide range of similarity index values indicated the presence of ample diversity among the genotypes analyzed. The greatest genetic diversity in this study was observed between ELG15–Selection3 and E77–EG45 (similarity value of 0.08) followed by Selection3–C43, BTx623–IS18551, ELG15–M35-1, ELG15–IS18551, E77–C43, E77–IS18551, E77–EG40, EG15–M35-1, EP 117–E145 and EJ67–IS8525 with a similarity value of 0.11. A dendrogram constructed based on the molecular marker polymorphism grouped the 24

sorghum genotypes into two major clusters with an average similarity of 24.4 % (Fig. 1). Cluster 1 included 11 genotypes that were grouped into four sub-clusters (1A, 1B, 1C and 1D). Sub-cluster 1A consisted of two genotypes (296B and B35) while sub-cluster 1B comprised three genotypes (CSV216R, EG40 and EG45). Three genotypes each were grouped in sub-clusters 1C (Selection3, M35-1 and IS18551) and 1D (EA11, EP117 and PEC17). Cluster 2 consisted of 13 genotypes that were grouped into three sub-clusters (2A, 2B and 2C). Sub-cluster 2A consisted of four genotypes (27B, B58586, IS8525 and E77) while sub-cluster 2C comprised two genotypes (EJ67 and UPMC503). Sub-cluster 2B was the major sub-cluster and comprised seven genotypes (C43, E36-1, SSV84, Btx623, EG15, ELG15 and E145). The majority of the genotypes belonging to post-rainy sorghum were contained in Cluster 1. Clusters 2 consisted of genotypes belonging to different races and there was no distinct grouping based on race. Forage sorghum genotypes (EJ67 and UPMC503) were grouped together in Cluster 2C. It is a well-known fact that most of the post-rainy sorghum varieties belong to *durra* type whereas *kharif* cultivars belong to *caudatum* and *kafir* races (Reddy et al. 2003). In addition, post-rainy sorghums are characterized by tolerance to shoot fly, stalk rot and terminal stress and possess large lustrous grain with semi-corneous endosperm (Reddy et al. 2006). It seems that the identification of clusters in this study is not based on sorghum races. Similarly, the study by Dje et al. (2000) pointed out that the races of sorghum are not substantially differentiated genetically. Moreover, Menz et al. (2004) and Gabriel (2005) reported that it is clear and evident that working groups and races in sorghum are not well defined. The main reason for not finding a clear clustering pattern among races could be the lack of representation of each race with a sufficient number of lines, as suggested earlier by Moncada (2006). Hence, it can be concluded that studies including a large number of lines per race and the use of greater numbers of marker loci should provide a better interpretation of genetic diversity in sorghum. However, the (GATA)_n motif-based microsatellite markers used in this study revealed the allelic richness among the diverse sorghum genotypes analyzed, thus making them suitable for application in crop genetics and marker-assisted breeding.

In conclusion, the present study was helpful in the identification of Class I (GATA)_n motif-containing microsatellites and their distribution pattern in the sorghum genome. Some of these motifs that are present in the genic region and upstream of genes might have functional significance, which needs to be investigated in detail. To the best of our knowledge, this is the first report of the genome-wide development and validation of (GATA)_n motif-based microsatellite markers in sorghum. The study also demonstrated the utility of these markers in genetic diversity analysis. Sharp and robust amplification combined with high polymorphism makes the (GATA)_n motif-based microsatellite markers suitable for genetic diversity assessment, cultivar identification, hybrid purity testing, QTL mapping and marker-assisted selection.

Acknowledgments The authors sincerely thank the Director, Directorate of Sorghum Research, Hyderabad, India, for the facilities and the encouragement provided for carrying out the study, and Dr. K. N. Ganapathy for his suggestions on improving the manuscript.

References

- Agrama HA, Tuinstra MR (2003) Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. *Afr J Biotechnol* 2:334–340
- Ali ML, Rajewski JF, Baenziger PS, Gill KS, Eskridge KM, Dweikat L (2008) Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. *Mol Breed* 21:497–509
- Areshchenkova T, Ganai MW (1999) Long tomato microsatellites are predominantly associated with centromeric regions. *Genome* 42:536–544
- Bout S, Vermerris W (2003) A candidate-gene approach to clone the sorghum brown midrib gene encoding caffeic acid O-methyltransferase. *Mol Gen Genomics* 269:205–214
- Chowdari KV, Venkatachalam SR, Davierwala AP, Gupta VS, Ranjekar PK, Govila OP (1998) Hybrid performance and genetic distance as revealed by the (GATA)₄ microsatellite and RAPD markers in pearl millet. *Theor Appl Genet* 97:163–169
- Danquah EY, Blay ET, Osae WM, Alifo Y (2000) Isozyme polymorphisms in eight accessions of sorghum (*Sorghum bicolor* L. Moench.). *J Ghana Sci Assoc* 2:93–98
- Davierwala AP, Ramakrishna W, Chowdari KV, Ranjekar PK, Gupta VS (2001) Potential of (GATA)_n microsatellites from rice for inter- and intra-specific variability studies. *BMC Evol Biol* 1:7
- Dice LR (1945) Measures of the amount of ecologic association between species. *Ecology* 26:297–302
- Dje Y, Ater M, Lefebvre C, Vekemans X (1998) Patterns of morphological and allozyme variation in sorghum

- landraces of North-western Morocco. *Genet Res Crop Evol* 45:541–548
- Dje Y, Heuertz M, Lefebvre C, Vekemans X (2000) Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. *Theor Appl Genet* 100:918–925
- Doggett H (1998) *Sorghum*, 2nd edn. Wiley, New York, p 512
- Fritz AK, Reddy AS, Pammi S, Ayres NM (1995) Silver staining as a low-cost, non-radioactive method of detecting PCR products. In: *Agronomy abstracts*, ASA, Madison, WI, p 184
- Gabriel K (2005) A study of heterotic relationships in sorghum. Ph.D. dissertation, Texas A&M University, College Station, TX, USA, p 99
- Grandillo S, Tanksley SD (1996) Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between *L. esculentum* and *L. pimpinellifolium*. *Theor Appl Genet* 92:957–965
- Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113: 163–185
- Harlan JR, De Wet MJM (1972) A simplified classification of cultivated sorghum. *Crop Sci* 12:172–176
- McCouch SR, Temnykh S, Lukashova A et al (2001) Microsatellite markers in rice: abundance, diversity and applications. In: Khush GS, Brar DS, Hardy B (eds) *Rice genetics IV*. IRRI, Manila, pp 117–135
- McNeil J, Smith KP, Hall LL, Lawrence JB (2008) Word frequency analysis reveals enrichment of dinucleotide repeats on the human X-chromosome and (GATA)_n in the X escape region. *Genome Res* 16:477–484
- Menz M, Klein R, Unruh N, Rooney W, Klein P, Mullet J (2004) Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. *Crop Sci* 44: 1236–1244
- Moncada RAF (2006) Genetic diversity and combining ability among sorghum conversion lines. Ph.D. dissertation, Office of Graduate Studies, Texas A&M University, College Station, TX, USA, p 108
- Mosges G, Friedtu W (1994) Genetic ‘Fingerprinting’ of sunflower lines and F₁ hybrids using isozymes, simple and repetitive sequences as hybridization probes, and random primers for PCR. *Plant Breed* 113:114–124
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76:5269–5273
- Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol Gen Genet* 252:597–607
- Parasnis AS, Ramakrishna W, Chowdari KV, Gupta VS, Ranjekar PK (1999) Microsatellite (GATA)_n reveals sex-specific differences in papaya. *Theor Appl Genet* 99:1047–1052
- Parh DK, Jordan DR, Aitken EAB, Mace ES, Junai P, McIntyre CL, Godwin ID (2008) QTL analysis of ergot resistance in sorghum. *Theor Appl Genet* 117:369–382
- Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457(7229):551–556
- Pena SDJ, Souza KTD, Andrade MD, Chakraborty R (1994) Allelic associations of two polymorphic microsatellites in intron 40 of the human *von Willebrand* factor gene. *Proc Natl Acad Sci USA* 91:723–727
- Rajendrakumar P, Biswal AK, Sakthivel K et al (2009) Development and validation of class I SSR markers targeting (GATA)_n repeat motifs in rice. *Euphytica* 169:263–271
- Rao R, Corrado G, Bianchi M, DiMauro A (2006) (GATA)₄ DNA fingerprinting identifies morphologically characterized ‘San Marzano’ tomato plants. *Plant Breed* 125:173–176
- Reddy BVS, Sanjana P, Ramaiah B (2003) Strategies for improving post-rainy season sorghum: a case study for landrace hybrid breeding approach. Paper presented in the workshop on heterosis in Guinea Sorghum, Sotuba, Mali, pp 10–14
- Reddy BVS, Ramesh S, Reddy PS (2006) Sorghum genetic resources, cytogenetics, and improvement. In: Singh RJ, Jauhar PP (eds) *Genetic resources, chromosome engineering and crop improvement*, Cereals, vol 2. CRC Press, Boca Raton, pp 309–363
- Riley DE, Krieger JN (2002) Diverse eukaryotic transcripts suggest short tandem repeats have cellular functions. *Biochem Biophys Res Commun* 298:581–586
- Rohlf FJ (2000) NTSYS-pc: numerical taxonomy and multivariate analysis system. Exeter Publishing, Setauket, p 31
- Saghai-Marouf MA, Soliman KM, Jordensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphism in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad Sci USA* 81: 8014–8018
- Scherzer CR, Grass JA, Liao Z, Pepivani I, Zheng B, Eklund AC, Ney PA, Ng J, McGoldrick M, Mollenhauer B, Brennick EH, Schlossmacher MG (2008) GATA transcription factors directly regulate the Parkinson’s disease-linked gene α -synuclein. *Proc Natl Acad Sci USA* 105(31): 10907–10912
- Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE, Paterson AH, Kresovich S (2002) Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench. *Theor Appl Genet* 105:912–920
- Serret MD, Udupa SM, Weigand F (1997) Assessment of genetic diversity of cultivated chickpea using microsatellite-derived RFLP markers: implications for origin. *Plant Breed* 116:573–578
- Shehzad T, Okuizumi H, Kawase M, Okuno K (2009) Development of SSR-based sorghum (*Sorghum bicolor* (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits. *Genet Resour Crop Evol* 56:809–827
- Singh L (1995) Biological significance of minisatellites. *Electrophoresis* 16:1586–1595
- Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. *Crop Sci* 40:226–232
- Subramanian S, Mishra RK, Singh L (2003) Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol* 4:R13

- Trembley JJ, Viger RS (2003) Transcription factor GATA-4 is activated by phosphorylation of serine 261 via the cAMP/Protein Kinase: a signaling pathway in gonadal cells. *J Biol Chem* 278:22128–22135
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93(1):77–78
- Wang ML, Zhu C, Barkley NA, Chen Z, Erpelding JE, Murray SC, Tuinstra MR, Tesso T, Pederson GA, Yu J (2009) Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theor Appl Genet* 120:13–23
- Wu Y, Huang Y (2008) Molecular mapping of QTL for resistance to the green bug *Schizaphis graminum* (Rondani) in *Sorghum bicolor* (Moench). *Theor Appl Genet* 117:117–124
- Yap IV, Nelson RJ (1996) WinBoot: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. IRRRI discussion paper series no. 14. International Rice Research Institute, P.O. Box 933, Manila, Philippines, p 22
- Yonemaru JI, Ando T, Mizubayashi T, Kasuga S, Matsumoto T, Yano M (2009) Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of sorghum (*Sorghum bicolor* (L.) Moench). *DNA Res* 16:187–193