Mol Breeding (2012) 29:543–564
DOI 10.1007/s11032-011-9571-0

Characterization, development and mapping of Unigene-derived microsatellite markers in sorghum [*Sorghum bicolor* (L.) Moench]

R. Nagaraja Reddy · R. Madhusudhana · S. Murali Mohan · D. V. N. Chakravarthi · N. Seetharama

Received: 23 October 2010 / Accepted: 21 March 2011 / Published online: 17 April 2011

Abstract Molecular variation within known genes controlling specific functions provide candidate gene-based markers which are tightly linked with the trait of interest. Unigene-derived microsatellite markers, with their unique identity and positions, offer the advantage of unraveling variation in the expressed component of the genome. We characterized ≥12-bp-long microsatellite loci from 13,899 unique sequences of sorghum [*Sorghum bicolor* (L.) Moench] available in the NCBI unigene database for their abundance and possible use in sorghum breeding. Analysis of 12,464 unigenes (≥200-bp) using MISA software identified 14,082 simple sequence repeats (SSRs) in 7,370 unigenes, from which 1,519 unigene SSR markers were developed. The average frequency of SSR was 1 per1.6 kb and 1.0 per 1.1 unigene; hexamers followed by trimers were found in abundance, of which 33.3% AT-rich and CCG repeats were the most abundant. Of the 302 unigene SSRs tested, 60 (19.8%) were polymorphic between the two parents, M35-1 and B35 of a recombinant inbred line (RIL) mapping population. A mapping population consisting of 500 RILs was developed using the above two parents, and a subset of random 245 RILs was used for genotyping with polymorphic SSRs. We developed a linkage map containing 231 markers, of which 228 (174 genomic and 54 genic) were microsatellites and three were morphological markers. Markers were distributed over 21 linkage groups, and spanned a genetic distance of 1235.5 cM. This map includes 81 new SSRs, of which 35 (21 unigene and 14 genomic) were developed in the present study and 46 from other studies. The order of the SSR markers mapped in the present study was confirmed physically by BLAST search against the whole-genome shotgun sequence of sorghum. Many unigene sequences used for marker development in this study include genes coding for important regulatory proteins and functional proteins that are involved in stress-related metabolism. The unigene SSR markers used together with other SSR markers to construct the sorghum genetic map will have applications in studies on comparative mapping, functional diversity analysis and association mapping, and for quantitative trait loci detection for drought and other agronomically important traits in sorghum.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-011-9571-0) contains supplementary material, which is available to authorized users.

R. Nagaraja Reddy · R. Madhusudhana (✉) · S. Murali Mohan · D. V. N. Chakravarthi · N. Seetharama
Marker Assisted Selection Lab, Directorate of Sorghum Research (DSR), Rajendranagar, Hyderabad 500 030, India
e-mail: madhu@sorghum.res.in

R. Nagaraja Reddy
Directorate of Medicinal and Aromatic Plants Research (DMAPR), Anand, Gujarat 387 310, India

S. Murali Mohan
International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Hyderabad 502 324, India

Keywords Sorghum · Unigene-derived microsatellite markers · Mapping · Drought tolerance

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is one of the most important cereal crops grown in semi-arid tropics of the world. It is traditionally grown as a food-fodder crop at subsistence levels by resource-poor farmers who cannot afford much input. In recent years, sorghum has emerged as a potential alternative bioenergy crop, thanks to tremendous demand for fuel-grade ethanol. The global area occupied by sorghum was about 46 Mha (FAOSTAT 2009) and leading producers of sorghum in the world include India, Nigeria, USA and Mexico. In India, sorghum is cultivated during both the *kharif* (rainy) and *rabi* (post-rainy) seasons. The area occupied by the crop during the fiscal year 2007–2008 was around 8.45 Mha, and the total production was 7.15 Mt with an average productivity of 846 kg/ha (FAOSTAT 2009). Sorghum breeding programmes across the world are working towards development of varieties with better quality, disease resistance, drought tolerance and agronomic traits (Klein et al. 2008).

In recent years, marker-assisted selection (MAS) has become popular in breeding programmes in many crops (e.g. William et al. 2007; Gupta et al. 2009; Ottoman et al. 2009). The development of high-density genetic linkage maps and subsequent quantitative trait locus (QTL) mapping of agronomically important traits is one of the prerequisites for employing MAS. In sorghum, many linkage maps have been developed using different marker types, including restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs) and Diversity Arrays Technology (DArTs) markers, and with major genes (Hulbert et al. 1990; Binelli et al. 1992; Whitkus et al. 1992; Berhan et al. 1993; Chittenden et al. 1994; Xu et al. 1994; Brown et al. 1996; Taramino et al. 1997; Boivin et al. 1999; Bhattramakki et al. 2000; Kong et al. 2000; Childs et al. 2001; Menz et al. 2002; Bowers et al. 2003; Wu and Haung 2006; Knoll et al. 2007; Ritter et al. 2008; Mace et al. 2008, 2009; Murray et al. 2008; Parh et al. 2008; Fernandez et al. 2008; Srinivas et al. 2008, 2009a, b; Satish et al. 2009; Ramu et al. 2009; Mace and Jordan 2010). Except for some markers from defined genes and/or expressed sequence tags (ESTs), many of the markers used in these linkage maps represent untranscribed regions of the genome which may not, in general, closely associate with variations in QTL or genes controlling phenotypic traits.

Microsatellites or SSRs are tandemly arranged repeats of DNA motifs 1–6 nucleotides long that frequently exhibit variation in the number of repeat units at a locus. SSRs are ubiquitous and are found in both protein coding and non-coding regions affecting gene expression (Cummings and Zoghbi 2000). They are favoured for a variety of applications in plant breeding because of their multi-allelic nature, reproducibility, co-dominant inheritance, high abundance and extensive genome coverage (Gupta and Varshney 2000). These markers are used in high-throughput genotyping and thus in the development of high-density genetic maps, gene mapping and MAS.

Although the utility of SSR markers has been well established, their de novo development can be costly and complex and is a time-consuming process on a locus-by-locus basis (Brown et al. 1996; Bhattramakki et al. 2000; Squirrell et al. 2003). Consequently, their development has been made much easier by the generation of large numbers of EST libraries (Cordeiro et al. 2001; Andersen and Lübberstedt 2003; Eujayl et al. 2004; Pratt et al. 2005; Srinivas et al. 2009a, b; Calviño et al. 2009) and by the availability of complete genome sequences (McCouch et al. 2002; Li et al. 2009) of main crop plants. EST databases are now available for more than 140 biologically and economically important plant species (Heesacker et al. 2008), and typically a large number of SSRs are abundant within these sequence collections (Ellis and Burke 2007). A major disadvantage of the EST-derived microsatellites is the sequence redundancy that yields multiple sets of markers at the same locus. However, SSRs developed from assembled ESTs or unigenes, popularly known as unigene-derived microsatellite (UGMS) markers, have the advantage of assaying variation in the expressed component of the genome with unique identity and positions (Parida et al. 2006). With the increasing emphasis on functional genomics, large number of EST sequences are being developed and assembled into unigene databases (http://www.ncbi.nlm.nih.gov/unigene) (Pontius et al. 2003, Lopez

et al. 2004). Furthermore, with evolving bioinformatics tools, it is now possible to identify and develop UGMS markers efficiently on a large scale in much less time (Parida et al. 2006). Because of the above advantages of UGMS markers, and relatively easy accessibility of large unigene resources, increasing numbers of UGMS markers are being identified and used for a variety of applications in a number of plant species like gerbera (Gong and Deng 2010), and cereals such as wheat, barley, maize, sorghum, rice and *Arabidopsis* (Parida et al. 2006). The UGMS markers can also be used for accurately assaying functional diversity in the natural populations and available germplasm collections as well as for comparative mapping and evolutionary studies as anchor markers.

Hence, in the present study, our main objective is to explore the unigene database of sorghum for mining genic SSRs, and to saturate our sorghum genetic linkage map with new SSR markers. Our long-term goal is to develop superior genotypes for resistance to drought and other stresses through genetic analysis, association studies and functional genomics studies.

Materials and methods

Development of mapping population

An F_8 recombinant inbred line (RIL) mapping population was developed consisting of 500 RILs derived from a cross using the sorghum parental lines, M35-1 and B35. M35-1 is a popular high-yielding post-rainy (*rabi)* variety of India famous for its pearly white grains having good *roti*-making qualities; whereas the parent B35 is a BC_1 derivative of IS12555, a *durra* race sorghum germplasm accession of Ethiopian origin. B35 is well characterized as a staygreen source conferring terminal drought tolerance and is used in QTL studies (Crasta et al. 1999; Xu et al. 2000; Subudhi et al. 2000; Sanchez et al. 2002; Harris et al.2007). From 500 RILs, 245 were randomly selected for genetic mapping studies. The two parents were hybridized during the *rabi* season of 2003 at the research farm of the Directorate of Sorghum Research (DSR), Hyderabad, India. A single F_1 plant yielding maximum seeds was advanced to develop F_2 generation seed. F_2 plants were grown and self-pollinated to obtain F_3 generations. $F_{2:8}$ generation RILs were developed following the single seed descent (SSD) method (Goulden, 1939).

Unigene database, marker searches and primer design

A total of 13,899 sequences of sorghum unigenes were obtained in bulk through a FASTA sequence retrieval system from GenBank using Batch Entrez (http://www.ncbi.nlm.nih.gov/) along with their constituent EST sequences (downloaded on March, 2009). The total size of the sequence was 8,735, 622 bp. These sequences were initially processed and mined for SSR motifs (monomers to hexamers) with a length of 12 bp and above for mono-, di-, tri-, tetra- and hexa-nucleotide repeats, and 15 bp and above for penta- nucleotide repeats, using a program MISA (MIcroSAtellite) written in the Perl 5 script language that locates microsatellite patterns in FASTA formatted sequence files and reports the GenBank ID, microsatellite motifs (monomers to hexamers), number of repeats and sequence coordinates for each microsatellite. The MISA script is available from the IPK website (http://pgrc.ipk-gatersleben.de/misa/), and is capable of identifying both perfect and compound microsatellites. The rational for choosing the small cutoff value was that the SSRs are often disrupted by single base substitutions (Subramanian et al. 2003). Microsatellites were classified into class I (>20 nucleotides), class II (12–20 nucleotides) and stochastic markers (class III, repeat length of 6–12 nucleotides) based on the length of the microsatellite motifs (Temnykh et al. 2000). SSRs with a motif length of 14 bp and above were selected for designing primers. Similarly, other genomic markers (Xdhsbm markers) were mined from the assembled whole-genome shotgun (WGS) sequence of sorghum available at http://www.phytozome.net/sorghum. These new SSR markers were developed by searching for di- to hexa-nucleotide repeats at every 50,000 bp across the genome. SSR with larger number of repeats was chosen for primer design if the unigene had more than one SSR. Primer pairs flanking SSRs were selected using Primer3 software (http://frodo.wi.mit.edu/primer3/). The key parameters set for primer design were as follows: primer length 18–24 bp with 20 bp as the optimum; PCR

product size 100–300 bp; optimum annealing temperature 54°C; GC content 35–60% with 50% as the optimum. The canonical name proposed for designating markers includes function [unigene (Un) or unknown (X)], lab designator (DSR Hyderabad (dh)], species [*Sorghum bicolor* (sb)], type of marker [microsatellite (m)] and serial no. of marker. Hence, the markers developed in this study were named 'Undhsbm' for unigenes and 'Xdhsbm' for markers from WGS. The primers were synthesized by Oscimum Biosolutions Pvt. Ltd, Hyderabad, India. The mapped EST genes or unigenes containing microsatellite motifs were searched against the predicted gene information of sorghum cultivar BTx623 using the BLASTX algorithm (http://www.phytozome.net/sorghum) for assigning putative functions.

DNA extraction, PCR amplification and electrophoresis

The genomic DNA was extracted using the CTAB method (Saghai-Maroof et al. 1984). PCR reactions were set up in a 5-μL reaction mixture in 384-well PCR plates (Axygen, PCR-384-HSC). Each PCR reaction mixture contained 2–4 pmol of primer, 1–4 mM $MgCl_2$, 0.1–0.2 mM dNTP, 0.2 U *Taq* DNA polymerase and 1× PCR buffer (Invitrogen, S. Giuliano Milanese, Italy) and 30–50 ng of genomic DNA as a template. Temperature cycling was carried out using the GeneAmp® PCR System9700 (Life Technologies Corporation, Carlsbad, CA, USA) and touch-down PCR amplification (Don et al. 1991): one 15-min denaturation cycle, followed by ten cycles of 94°C for 10 s, 61°C for 20 s (reducing by 1°C per cycle) and 72°C for 30 s, then by 31 cycles of 94°C for 10 s, 54°C for 20 s and 72°C for 30 s. After completion of the 31 cycles, a final extension of 20 min at 72°C was included to minimize the +A overhang (Smith et al. 2000). PCR products were separated on 4% agarose (Sigma–Aldrich, St. Louis, MO, USA) gels as well as on sequencing electrophoresis apparatus (Bio-Rad Laboratories, Philadelphia, PA, USA) in 6% polyacrylamide gel containing 8 M urea and 1 × TAE buffer at 80 W of constant power. The DNA fragments were visualized by silver staining (Fritz et al. 1999) and scored either as 'A' (M35-1 allele) or 'B' (B35 allele), heterozygous (H), or missing data (–).

Genetic mapping

Genomic DNA of M35-1 (parent 1) and B35 (parent 2) were subjected to PCR amplification by using both new and published SSR markers. The polymorphic markers were used for genotyping 245 F_8 RILs. The software Joinmap 3.0 (Van Ooijen and Voorrips 2001) was used for linkage map construction. The allocation of markers to linkage groups was mostly stable for a wide range of LOD grouping thresholds (from $\leq$4.0 to $\geq$7.0). More stringent parameter settings were not applied where the marker location were already known from the earlier published reports. The Kosambi mapping function was used to convert recombination into genetic distance (centimorgans, cM). The goodness-of-fit of the constructed maps reflecting the discrepancy between final recombination frequencies in the map and those apparent from individual marker data pairs was expressed as a chi-squared value and computed according to Stam and van Ooijen (1995). Linkage groups were named as proposed by Kim et al. (2005). The distance measurement of interval variables between two individuals, proposed originally by Gower et al. (1971) and modified by Wu and Haung (2006), was used to compare the genetic distances between the present map and the maps of Srinivas et al. (2009b) and Wu and Haung (2006) and the consensus map of Mace et al. (2009). The modified distance measure (Wu and Haung 2006) is based on the following formula: difference ratio $= \Sigma\ |A_{ik} - B_{ik}|/(A_i + B_i)$, where A_{ik} is the length (in cM) of the *k*th shared marker interval on the *i*th chromosome of map A, B_{ik} is the length (in cM) of the *k*th shared marker interval on the *i*th chromosome of map B, $\Sigma|A_{ik} - B_{ik}|$ is the absolute value of the length difference of each shared marker interval on the *i*th chromosome between maps A and B and $A_i + B_i$ is the additive value of all shared intervals for the *i*th chromosome of maps A and B which is used to normalize the difference value, $\Sigma|A_{ik} - B_{ik}|$. The RIL population was also evaluated for segregation of three morphological markers, viz., seed pericarp colour (*Pericarp)*, plant colour intensity (*PlcorInt)* and yellow midrib colour (*Ymrco*). The expression of plant colour intensity was scored as 'A' for light red colour of parent M35-1 and 'B' for dark brown red colour of parent B35; pericarp colour as 'A' for yellowish white seed of M35-1 and 'B' for brownish pericarp of B35, and yellow midrib colour

as 'A' for white midrib of M35-1 and 'B' for yellow midrib of B35, at the time of physiological maturity, and used as morphological markers in the linkage analysis. This experiment was repeated for three post-rainy seasons from 2006 to 2008 to confirm the segregation of these morphological markers.

Physical mapping of SSR markers

For each SSR marker, the forward and reverse primers were compiled and the primers were BLAST searched (http://www.phytozome.net/search.php?show=blast&blastdb=sorghum1) against aligned WGS information of the elite sorghum inbred line BTx623 (Paterson et al. 2009), and their physical positions were recorded. In cases where the primer pairs did not give a match with the genome sequence, the corresponding source sequence (probe/EST), if available in the public domain, was used for determining their physical positions. Markers with no BLAST hits and source sequence data were not included in the physical map. An E-value of zero or close to zero was adopted to claim a significant match between primer and the sorghum genome. The BLAST results were further analyzed using the conditions suggested by Ramu et al. (2009). Wherever multiple BLAST hits were found for genomic and genic markers other than unigene markers, linkage map information was used in order to determine the position. The primers satisfying the above criteria were placed on the sequence-based physical map of the sorghum genome. MapChart 2.1 software (Voorrips 2002) was utilized for physical map construction.

Results

Genic-SSR frequency, distribution and marker development

A total of 13,899 sorghum unigenes were used in the study, of which 12,464 were selected after initial processing and exclusion of sequences smaller than 200 bp for SSR search. These sequences represented 8,735,622 bp of the putative functional transcriptome. The MISA-based microsatellite search of these unigenes detected a total of 14,082 SSRs in 7,370 sequences (59.1% of the 12,464 unigenes), suggesting an average frequency of 1 SSR per 1.6 kb and 1.0 per 1.1 unigene of the sorghum transcriptome analyzed. Analysis of SSR motifs in these sequences revealed 3,621 (49.1%) unigenes had more than one SSR and 282 (2%) had SSR motifs in non-interrupting compound formation. The frequency of SSR motifs revealed that hexamers are more frequent (63.1%) (Table 1, Electronic Suplementary Material Fig. 1), followed by trimers (20%), tetramers (6.6%), monomers (6.0%), dimers (2.5%) and pentamers (1.9%). In general, AT-rich motifs were found to be more frequent in all types of repeat motifs except for trimers where CG-rich motifs were predominantly observed. Among hexamers, 33.3%AT-rich (27%) motifs were most common followed by 50%AT-rich (25%) motifs and 16.6%AT-rich (22%) motifs, whereas 66.6%AT-rich (13.8%), 0%AT-rich (5.6%), 83.3%AT-rich (5.4%) and 100%AT-rich (0.6%) motifs were found less frequently. Similarly, among the trimers, the motif CCG was most common (41.4%), followed by AGC (13.9%) and AGG (12.2%), whereas the motif AAT was least common (1.0%). However, monomers, dimers, tetramers, or pentamers were found in insignificant numbers (<10%). Of the total number of SSRs studied, 249 (1.7%) were found to be hypervariable (class I, repeat length of ≥20 bp), of which 242 were monomeric and 7 dimeric SSRs; 2,499 (17.7%) SSRs were potentially variable markers (class II, repeat length of 13–20 bp), of which 545, 211, 1054, 178, 66 and 445, respectively were from mono- to hexamers, and the remaining were stochastic markers (class III, repeat length of 6–12 bp). A total of 1,519 SSR-unigene sequences containing SSRs of length ≥14 bp and di- to hexamers were used for primer design. The monomers were excluded from the final list since practical problems related to allele sizing were expected. Of these, primers were synthesized for 313 unigene SSRs containing 292 perfect and 19 compound SSRs (Supplementary Table 1).

Detection of parental polymorphism

The genomic DNA of the two parents M35-1 and B35 were screened for polymorphism using 313 unigene markers (Supplementary Table 1) developed in the present study along with 762 SSR markers reported from previous studies (Bhattramakki et al. 2000; Schloss et al. 2002; Srinivas et al. 2008 and 2009a; Ramu et al. 2009). Of the 313 unigene primer pairs, 302 primer pairs produced amplification products that

Table 1 Distribution of SSRs in unigene database of sorghum

SSR type	Motif type	Motif length (repeat units)																			Total	Motif (%)	Total (%)
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	≥20			
Mono	A/T											57	33	32	37	43	166	147	76	238	829	98.1	
	C/G											1	4	4	1	1	1	–	–	4	16	1.9	
	Total											**58**	**37**	**36**	**38**	**44**	**167**	**147**	**76**	**242**	**845**		**6.0**
Di	AC/GT					39	12	8	5	7	7	1	2	2	1	1	1	–	–	1	87	24.9	
	AG/CT					46	30	26	16	10	6	5	6	5	2	5	3	–	1	5	166	47.4	
	AT/AT					31	9	7	4	4	2	3	3	1	–	–	–	1	–	1	66	18.9	
	CG/CG					16	10	5	–	–	–	–	–	–	–	–	–	–	–	–	31	8.9	
	Total					**132**	**61**	**46**	**25**	**21**	**15**	**9**	**11**	**8**	**3**	**6**	**4**	**1**	**1**	**7**	**350**		**2.5**
Tri	AAC/GTT			39	15	4	1	2	1	–	–	–	–	–	–	–	–	–	–	–	62	2.2	
	AAG/CTT			108	36	7	2	1	1	1	1	–	–	–	–	1	–	–	–	–	158	5.6	
	AAT/ATT			19	6	3	1	–	–	–	–	–	–	1	–	–	–	–	–	–	30	1.1	
	ACC/GGT			135	28	23	6	–	1	1	–	–	–	–	–	–	–	–	–	–	194	6.9	
	ACG/CTG			212	67	33	15	3	1	–	–	–	–	–	–	–	–	–	–	–	331	11.8	
	ACT/ATG			63	15	9	2	1	1	1	–	–	–	–	–	–	–	–	–	–	92	3.3	
	AGC/CGT			226	86	45	20	8	4	1	–	–	–	–	1	—	–	–	–	–	391	13.9	
	AGG/CCT			221	66	36	13	5	2	1	–	–	–	–	–	–	–	–	–	–	344	12.2	
	AGT/ATC			34	7	2	2	–	–	–	–	–	–	–	–	–	–	–	–	–	45	1.6	
	CCG/CGG			701	297	100	38	20	5	3	–	–	1	–	–	–	–	–	–	–	1165	41.4	
	Total			**1758**	**623**	**262**	**100**	**40**	**16**	**8**	**1**	–	**1**	**1**	**1**	**1**	–	–	–	–	**2812**		**20.0**
Tetra	0%AT		83	11	3	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	98	10.6	
	25%AT		166	38	13	3	1	–	–	–	–	–	–	–	–	—	–	–	–	–	221	23.9	
	50%AT		334	50	17	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	402	43.5	
	75%AT		151	23	7	4	2	–	1	–	–	–	1	–	–	1	—	–	–	–	190	20.5	
	100%AT		13	–	–	1	–	–	–	–	–	–	–	—	–	—	–	–	–	–	14	1.5	
	Total		**747**	**122**	**40**	**9**	**3**	–	**1**	**1**	–	–	**1**	–	–	**1**	–	–	–	–	**925**		**6.6**
Penta	0%AT		15	2	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	18	6.8	
	20%AT		39	10	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	50	18.9	
	40%AT		72	22	8	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	102	38.5	
	60%AT		51	10	3	1	–	1	–	–	1	–	–	–	–	–	–	–	–	–	67	25.3	
	80%AT		22	5	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	28	10.6	
	100%AT		–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0.0	

Table 1 continued

SSR type	Motif type	Motif length (repeat units)																		Total	Motif (%)	Total (%)
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	≥20		
	Total		**199**	**49**	**14**	**1**	–	**1**	–	–	**1**	–	–	–	–	–	–	–	–	–	**265**		**1.9**
Hexa	0%AT	450	45	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	497	5.6	
	16.6%AT	1780	140	31	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1954	22.0	
	33.3%AT	2267	115	15	2	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2400	27.0	
	50%AT	2212	53	9	1	—	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2275	25.6	
	66.6%AT	1209	17	2	0	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	1229	13.8	
	83.3%AT	466	9	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	477	5.4	
	100%AT	52	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	53	0.6	
	Total	**8436**	**380**	**60**	**7**	**1**	–	**1**	–	–	–	–	–	–	–	–	–	–	–	–	**8885**		**63.1**
	Total	8436	1326	1989	684	405	164	88	42	30	17	67	50	45	42	52	171	148	77	249	14082		

were single and easy to score, while the remaining 11 (3.5%) produced either multiple fragments that were difficult to score or showed no amplification. Out of 302 locus-specific primer pairs, 60 genic-SSRs (19.8%) were found to be polymorphic between the parental lines with expected amplicon sizes. In the present study, a set of 33 of the 60 polymorphic unigene-SSR markers were used for genotyping the 245 RILs and for linkage map construction.

Linkage mapping

Altogether, 276 SSR markers consisting of 210 genomic-SSRs and 60 genic-SSRs including 33 unigenes and three morphological markers were used for construction of the linkage map. Sixty markers out of 276 polymorphic markers (21.7%) screened in the mapping population showed segregation distortion ($P < 0.1$) from the expected Mendelian ratio (1:1) and were skewed towards female parent. A significant number of the markers showing distorted segregation were also linked by at least 5 cM to markers that did not show distorted segregation patterns. Of these deviated markers ($P < 0.01$), seven loci (Xisep1038, Xgap206, Xtxp43, Xtxp329, Gpsb89, Xtxp88 and Xisep1028) on SBI-01, and 20 loci (Xdhsbm1164, Xdhsbm1157, Undhsbm66, Xisep948, Xisep224, Gpsb151, Xdhsbm1159, Xdhsbm1173, Xdhsbm1194, Xdhsbm1189, Xdhsbm1172, Xdhsbm1168, Xcup48, Xdhsbm1182, Xdhsbm1195, Xcup05, Xcup23, Xisp230, Gpsb50 and Xtxp26) on SBI-04 were clustered and mapped as blocks on the chromosome. Likewise, four loci (Xisp326, Xisep604, Undhsbm217 and Undhsbm105) on SBI-10-1, and ten loci (Xtxp290, Xtxp353, Xtxp337, Xtxp20, Xtx331, Xtxp270, Xisp359, msbCIR283, Xisep625 and Xsbarslbk10.09) on SBI-10-2 were clustered. All the deviated markers were retained in the linkage analysis since their map positions were consistent with the earlier reports (Bhattramakki et al. 2000; Srinivas et al. 2008, 2009a, b) and also with the in-silico map in the present study. A new sorghum genetic map comprising 231 loci with 228 SSR markers (174 genomic and 54 genic) and three morphological markers was constructed. Of the 228 SSR markers, 81 were novel SSRs and the remaining were mapped in other studies. Forty-six marker loci including 12 unigene markers were unlinked or greatly disturbed the map construction, thereby decreasing the goodness-of-fit of the map.

These markers were removed from the data set. Thirty-five out of 81 new SSR markers mapped were developed in the present study and the remaining 46 markers had been reported earlier (Brown et al.1996; Kong et al. 2000; Bhattramakki et al. 2000; Schloss et al.2002; Srinivas et al. 2008, 2009a; Ramu et al. 2009) but their map positions were not known. The primer sequence information and SSR motifs of the new SSRs are given in Table 2. Five unigene markers each were mapped to SBI-01 and SBI-03, three each on SBI-02, SBI-04 and SBI-10, and one each on SBI-08 and SBI-09 (Table 2; Fig. 1).

According to linkage analysis, 21 linkage groups were identified based on published anchor SSRs (Bhattramakki et al. 2000; Wu and Haung 2006; Srinivas et al. 2009b) and the in-silico map of published SSR markers (present study). Two chromosomes (SBI-03 and SBI-05) had one linkage group each, whereas each of the eight other chromosomes showed two or three linkage fragments. The map covered 1235.5 cM, with an average distance of 6.64 cM between the markers. The number of markers per linkage group ranged from seven (SBI-07) to 45 (SBI-01) and the length of each linkage group of the chromosomes varied from 36.4 cM (SBI-09) to 172.7 cM (SBI-01) (Fig. 1). Large gaps (>20 cM) were found on linkage groups SBI-05 (between markers Xdhsbm1076 and Gpsb17), SBI-06 (between markers Xtxp127 and Xtxp145, Xisep443 and Xisep427), SBI-07-2 (msbCIR300–Xtxp295) and SBI-08 (Xtxp273–Drenhsbm16). On SBI-06, in the gap between markers Xtxp127 and Xtxp145, the unigene polymorphic marker Undhsbm95 is found; similarly, in the gap between markers Xisep443 and Xisep427, the polymorphic unigene markers Undhsbm99 and Undhsbm76 are found. In the gap between msbCIR300 and Xtxp295 on SBI-07-2, the polymorphic marker Undhsbm60 is found. However, on SBI-05 no unigene marker was found to be polymorphic. Thus, with the use of new unigenes which are polymorphic between the parents, more saturation can be achieved in the large gaps.

Linkage groups SBI-01-1 and SBI-01-2 of chromosome SBI-01 together spanned a distance of 172.7 cM, which was the longest. This linkage group consists of 45 markers including 17 new markers (Fig. 1) and a morphological marker. Thirty-two SSRs were mapped in three linkage groups (SBI-02-1, SBI-02-2 and SBI-02-3) of chromosome SBI-02, including six new loci, which covered a distance of 169.7 cM. The linkage group corresponding to SBI-03 contains 35 SSR loci with 11 loci newly assigned and covered a distance of 155.6 cM. Linkage groups SBI-04-1, SBI-04-2 and SBI-04-3 together covered a distance of 168.8 cM and includes 31 SSRs, with 17 new SSR loci and a morphological marker (*PlcorInt*). Similarly, SBI-05, SBI-06, SBI-07, SBI-08, SBI-09 and SBI-10 covered respective distances of 80.6, 126.8, 109.7, 120.8, 36.4 and 97.4 cM, and were mapped with 7 (5 new), 16 (4 new), 15 (5 new), 16 (2 new) 9 (4 new) and 22 (10 new) markers, respectively. The extent of representation of the present linkage map with that of the recently published consensus genetic map containing DArT, SSR and AFLP markers (Mace et al. 2009) was estimated (Supplementary Table 2). Upon comparison, it was found that the extent of chromosome representation in our study varied from 20.7% (SBI-09) to 99.6% (SBI-04). Across the genome, our map represented about 77.3% of the consensus map and there were 76 markers common between the two linkage maps (data not provided). Good colinearity of the marker order was observed along most of the linkage groups.

A difference ratio as suggested by Wu and Haung (2006) was calculated per chromosome to compare the genetic distance between the genetic map of the present study with recent SSR maps of Wu and Haung (2006) and Srinivas et al. (2009b) and the consensus map of DArT, RFLP, AFLP and SSRs markers of Mace et al. (2009). The difference ratios calculated in each population are presented in Supplementary Table 3, where a difference ratio of "0" indicates identical genetic distance between the two maps and a difference ratio of "1" indicate complete dissimilarity of genetic distance between two maps. The number of intervals in common between the map of the present study with other maps varied; it was 30, 29 and 55 intervals with the maps of Wu and Haung (2006) Srinivas et al. (2009b) and Mace et al. (2009), respectively. The overall difference ratios in genetic distance between the map of the present study with other maps were 0.13 (Wu and Haung 2006), 0.19 (Srinivas et al. 2009a, b) and 0.10 (Mace et al. 2009), indicating good similarity of genetic distance.

Mapping of morphological markers

Three morphological markers, pericarp colour (*Pericarp*; brown/white), plant colour intensity (*PlcorInt*;

Table 2 Details of 35 new markers developed and mapped in this study

S. No.	Marker	Unigene ID	Repeat motif	Reverse primer sequence (5′–3′)	Reverse primer sequence (3′–5′)	T_m (°C)	Approx. amplicon size (bp)	Linkage group	Physical map position
1	Undhsbm1	Sbi.63	$(CGA)_7$	GCCCACCAAATACAAGAACC	CTTTGCGGCACTAAAACACA	54	236	SBI-01	SBI-01
2	Undhsbm24	Sbi.736	$(GCA)_{10}$	GCAATGGAGTCTTGGACGAT	GAAACCAGAACCGAACCTGA	60	173	SBI-01	SBI-01
3	Undhsbm68	Sbi.2622	$(TC)_{16}$	CTACTGGCTCATTGCTTCTCTCTC	AAAATGGGACCTCGAAAACCT	61	173	SBI-01	SBI-01
4	Undhsbm90	Sbi.3399	$(GCA)_7$	CAAGGAGCTGTGGAAGAAGG	CCTCCTGCTCCTGTGGTAAC	60	178	SBI-01	SBI-01
5	Undhsbm195	Sbi.8865	$(GGC)_7$	AGAGACCAACAGGCCTCTCC	CGTGTACTCGCGGATGTAGT	54	180	SBI-01	SBI-01
6	**Undhsbm52**	**Sbi.2012**	**$(TAG)_7$**	**CTTCATTAACCGCGACGACT**	**CTTCCCACGCACAGTAATCA**	**60**	**204**	**SBI-03**	**SBI-02**
7	Undhsbm82	Sbi.3125	$(CTT)_{10}$	ATCAAGCAGCAGAGGAGCAT	GCTTGATGAGCTCCCTTACG	60	247	SBI-02	SBI-02
8	Undhsbm154	Sbi.7032	$(AGC)_{15}$	CAGAGCATACCTCCCCTGAA	TTGAATCGGTTGCATGGATA	54	239	SBI-02	SBI-02
9	Undhsbm159	Sbi.7366	$(AC)_{11}$	GAATGATCACGAGGCTCCAT	CATGCCTCCCTTCTCTGAAC	54	154	SBI-02	SBI-02
10	Undhsbm56	Sbi.2161	$(TG)_{16}$	GCTGTGAATTGCTTCCTATCG	CCTCTGAGTTTTCACACACACAC	60	212	SBI-03	SBI-03
11	Unnhsbm314	Sbi.10880	$(AGG)_{12}$	TTTCGAGATCTCCACGACCT	CAGTGCTAACTCCCGATGCT	60	190	SBI-03	SBI-03
12	Undhsbm164	Sbi.7738	$(AC)_{14}$	ATACCTGCCTCCCTCTGGAT	TCGGTTCTTGCTCAACACCT	54	164	SBI-03	SBI-03
13	Undhsbm209	Sbi.9424	$(TC)_{10}$	GTTACGGGTGATTCTTCTCG	GATCAGCGCGATTGCTTC	53	167	SBI-03	SBI-03
14	Undhsbm66	Sbi.2553	$(TGC)_7$	ATACTGCCCACTTGGTTTCG	GGCAACAAGCAGTTGGAAAT	60	218	SBI-04	SBI-04
15	Undhsbm210	Sbi.9457	$(AT)_{13}$	CGGCCAGAGTAACTTTCCAC	CAACTGCCATTTCTTCAGCA	54	222	SBI-04	SBI-04
16	Undhsbm265	Sbi.11534	$(CTT)_9$	TGCACAAGAGGCAGGATATG	AGCCCATGAGAAAAAGCTCA	54	233	SBI-04	SBI-04
17	Undhsbm109	Sbi.4081	$(AG)_7$	AAGTCCTCCCCCGTATGTTC	TCACAATACAGGCAGGCAAG	60	196	SBI-08	SBI-08
18	Undhsbm178	Sbi.8301	$(AT)_9$	TAACTTGGCAAGACGGGATG	GTCGGTTCCGAGGGTTACTA	54	178	SBI-09	SBI-09
19	Undhsbm105	Sbi.4002	$(GA)_{13}$	CCATCCCACAACGAAGAAAC	GACGTCCCAGTCGAAGAATG	61	186	SBI-10	SBI-10
20	Undhsbm217	Sbi.9792	$(TC)_{12}$	ACCACTACCTGGGGTCTGTG	GAAATTCATGAACCCCAACG	54	230	SBI-10	SBI-10
21	Undhsbm313	Sbi.17757	$(TAAA)_{13}$	CAGCTACGATTTGGGGAAAA	CCTCAACCAAATTGCACAAA	55	170	SBI-10	SBI-10
22	Xdhsbm1005		$(AGATAT)_{16}$	TGGACTAGCGAGGCAGATAAC	CTGCGAATGTGACTTTTCTGA	55	214	SBI-10	SBI-10
23	Xdhsbm1076		$(AATA)_7$	AAAAGATAGTCCATAAAGGGCTTACA	CTGTAACAAATAAAGTGAAACAAACG	55	250	SBI-05	SBI-05
24	Xdhsbm1157		$(TAAA)_6$	GCTCCCAAATCTCGATAACAA	ACTGGTTTAGGCGGATCCTT	55	231	SBI-04	SBI-04
25	Xdhsbm1159		$(AT)_{36}$	TGAGGTGATTGTCTGGGTGA	GCCTGCACAAAGAAACAAAA	55	231	SBI-04	SBI-04
26	Xdhsbm1164		$(AG)_{23}$	AACCGTGCCTGATGCTAATC	GCTCCAACCACAGAAACCAT	55	223	SBI-04	SBI-04
27	Xdhsbm1168		$(GTC)_7$	ACAACATGTGACAGCCGGTA	TCTGGGTAGGTGAGCGAATC	55	198	SBI-04	SBI-04
28	Xdhsbm1172		$(CGC)_7$	TTTTCCCAGCACGAAATTCT	TTAGGACTCGCAAGGAGGAG	55	215	SBI-04	SBI-04
29	Xdhsbm1173		$(CTCCG)_4$	GAGATCCAGTCGCTGCTAGG	GAACGAACGAACCACCACTT	55	230	SBI-04	SBI-04
30	Xdhsbm1182		$(CA)_{12}$	TCGCCCTACATGTCATACCA	ACTCCATTGAGGGTGAGGTG	55	200	SBI-04	SBI-04
31	Xdhsbm1189		$(AAG)_{10}$	CCCACCATAAACATTTCACCA	AACCGATCGTTTTACCTCTTGA	55	232	SBI-04	SBI-04

Table 2 continued

S. No.	Marker	Unigene ID	Repeat motif	Reverse primer sequence (5′–3′)	Reverse primer sequence (3′–5′)	T_m (°C)	Approx. amplicon size (bp)	Linkage group	Physical map position
32	Xdhsbm1194		$(TCGA)_5$	TTGGGGACGACTAGAGGTTG	TTCAACGTCACCAGGTCAGA	55	224	SBI-04	SBI-04
33	Xdhsbm1195		$(CGG)_9$	CTAAAGGAACTCGGCGATTG	GTCGTGTCCTTCGGCATTAT	55	174	SBI-04	SBI-04
34	Xdnhsbm10		$(GCGATG)_4$	TATCTCATCCACGGAAAGC	GTCCACTCCAACCACAACG	60	205	SBI-09	SBI-09
35	Xdnhsbm43		$(CAGAG)_5$	GAAAACCCAATTGACCTCCA	TGTCTCGTACTTGATGGTCTGC	60	197	SBI-04	SBI-04

S. Nos. 1–21 were developed using the unigene database of sorghum and the remainder were developed from the whole-genome sequence of sorghum; markers conflicting in their physical and genetic map positions are highlighted in bold

Fig. 1 Alignment of genetic map (cM) of M35-1 × B35 population (*left*) with in-silico map (Mb; *right*) of SSR markers mapped in this study. Linkage group nomenclature follows Kim et al. (2005). Newly added markers are in bold and genic markers are underlined. Only genetic and physical map positions of newly mapped markers are represented with *lines*. *Black bars* indicate the regions of segregation distortion. ***Pericarp**** on SBI-01-2, ***PlcorInt**** on SBI-04-2 and ***Ymrco**** on SBI-06-1 are the three morphological markers. # Indicates markers that did not produce BLAST hits with sorghum genome. ** Indicates markers that are conflict in their physical and genetic map positions

light red/dark red) and yellow midrib colour (*Ymrco*; yellow/white), were mapped in this population. *Pericarp* locus was mapped on SBI-01 flanked by Xtxp320 and Undhsbm195, *PlcorInt* locus was mapped on SBI-04 between Undhsbm210 and Xisp343 and *Ymrco* locus was mapped on SBI-06 flanked by Xcup12 and Xcup37.

In-silico mapping

The recently published sorghum genome sequence (Paterson et al. 2009) has facilitated comparison of the position and order of genetically mapped markers in the present genetic linkage map with that of the physical map. The map order of the genetically mapped SSR markers in the present study was confirmed physically by BLAST search against the WGS of sorghum. Most of the genetically mapped markers were in good agreement with the physical map positions and order (Fig. 1). However, the markers Xtxp208, Xtxp329, Xisp324, Xtxp216 (genetically mapped to SBI-01), SbAGH04, Xtxp304 (SBI-02), Xtxp267, Xtxp218 (mapped to SBI-03), Gpsb50, Xisp343 (SBI-04), Gpsb127 and Gpsb69 (SBI-06), Xgap15 (SBI-09), Xtxp290, Xtxp20, Xtxp270, Xisp359 and Xsbarslbk10.09 (SBI-10) produced no hits with the genome sequence and were not included in the physical map. The markers were then aligned on the sorghum physical map (Fig. 1) using Mapchart 2.1 software. The in-silico map thus contains 209 markers. The physical map positions and order of markers were essentially as expected from the genetic linkage analysis of these markers, except for the markers Xcup61, Undhsbm52, Xdnhsbm43, Xtxp353 and Xgap05 which were genetically mapped to SBI-01, SBI-03, SBI-04, SBI-07 and SBI-08 linkage groups, but physically mapped to chromosome SBI-03, SBI-02, SBI-05, SBI-10 and SBI-09, respectively.

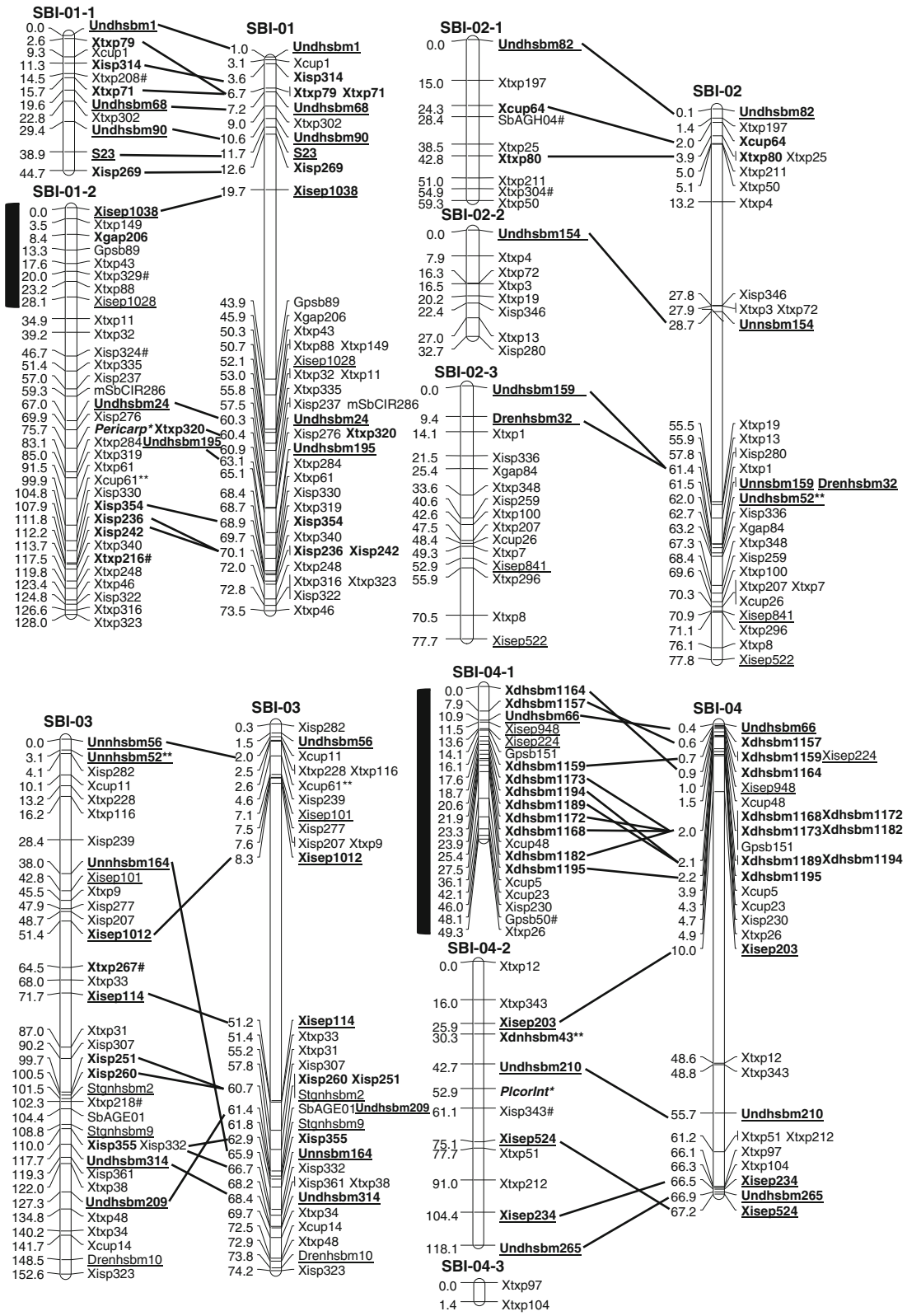
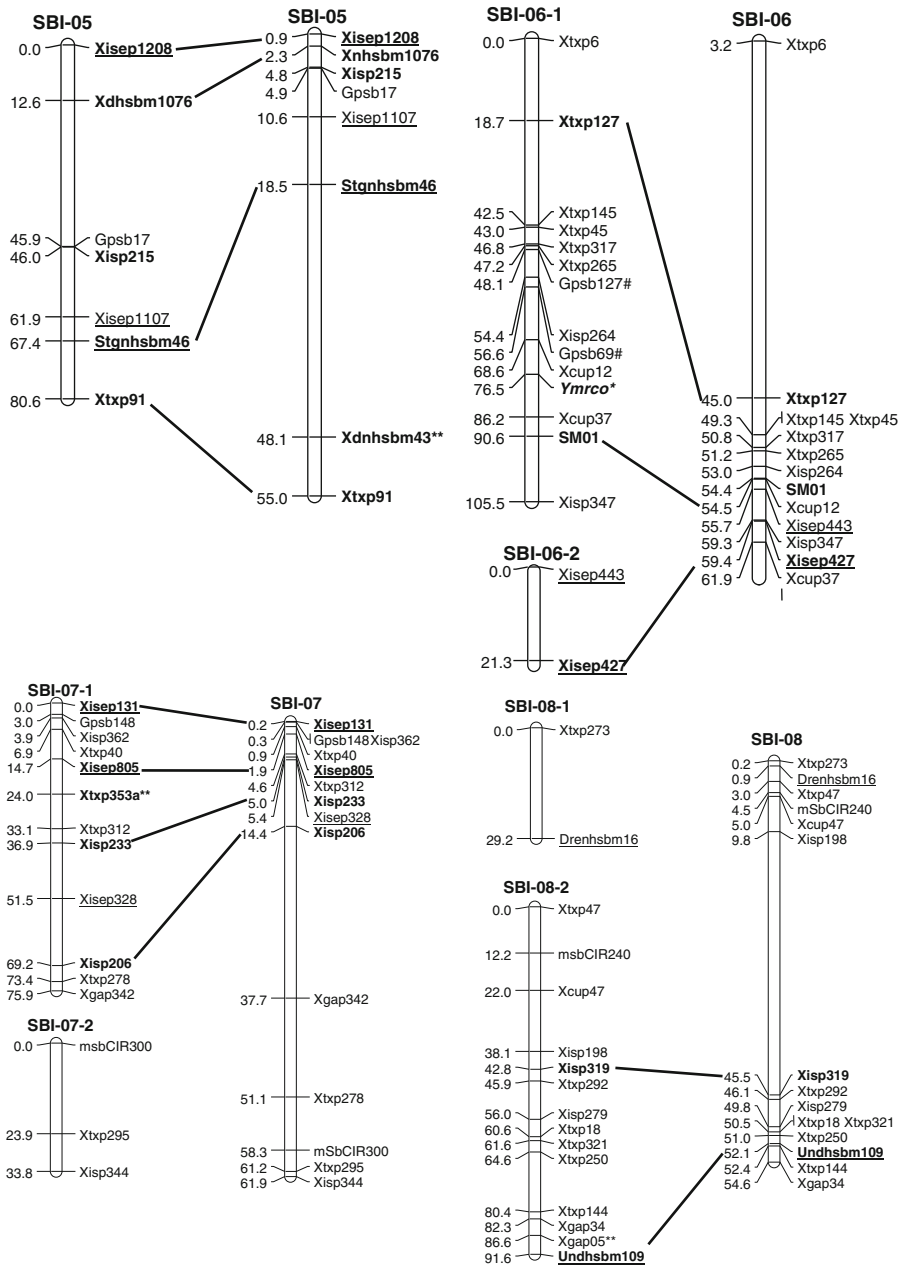

SBI-01-1
0.0 Undhsbm1
2.6 Xtxp79
9.3 Xcup1
11.3 Xisp314
14.5 Xtxp208#
15.7 Xtxp71
19.6 Undhsbm68
22.8 Xtxp302
29.4 Undhsbm90
38.9 S23
44.7 Xisp269
SBI-01-2
0.0 Xisep1038
3.5 Xtxp149
8.4 Xgap206
13.3 Gpsb89
17.6 Xtxp43
20.0 Xtxp329#
23.2 Xtxp88
28.1 Xisep1028
34.9 Xtxp11
39.2 Xtxp32
46.7 Xisp324#
51.4 Xtxp335
57.0 Xisp237
59.3 mSbCIR286
67.0 Undhsbm24
69.9 Xisp276
75.7 Pericarp* Xtxp320
83.1 Xtxp284 Undhsbm195
85.0 Xtxp319
91.5 Xtxp61
99.9 Xcup61**
104.8 Xisp330
107.9 Xisp354
111.8 Xisp236
112.2 Xisp242
113.7 Xtxp340
117.5 Xtxp216#
119.8 Xtxp248
123.4 Xtxp46
124.8 Xisp322
126.6 Xtxp316
128.0 Xtxp323
SBI-01
1.0 Undhsbm1
3.1 Xcup1
3.6 Xisp314
6.7 Xtxp79 Xtxp71
7.2 Undhsbm68
9.0 Xtxp302
10.6 Undhsbm90
11.7 S23
12.6 Xisp269
19.7 Xisep1038
43.9 Gpsb89
45.9 Xgap206
50.3 Xtxp43
50.7 Xtxp88 Xtxp149
52.1 Xisep1028
53.0 Xtxp32 Xtxp11
55.8 Xtxp335
57.5 Xisp237 mSbCIR286
60.3 Undhsbm24
60.4 Xisp276 Xtxp320
60.9 Undhsbm195
63.1 Xtxp284
65.1 Xtxp61
68.4 Xisp330
68.7 Xtxp319
68.9 Xisp354
69.7 Xtxp340
70.1 Xisp236 Xisp242
72.0 Xtxp248
72.8 Xtxp316 Xtxp323
73.5 Xisp322
Xtxp46
SBI-02-1
0.0 Undhsbm82
15.0 Xtxp197
24.3 Xcup64
28.4 SbAGH04#
38.5 Xtxp25
42.8 Xtxp80
51.0 Xtxp211
54.9 Xtxp304#
59.3 Xtxp50
SBI-02-2
0.0 Undhsbm154
7.9 Xtxp4
16.3 Xtxp72
16.5 Xtxp3
20.2 Xtxp19
22.4 Xisp346
27.0 Xtxp13
32.7 Xisp280
SBI-02-3
0.0 Undhsbm159
9.4 Drenhsbm32
14.1 Xtxp1
21.5 Xisp336
25.4 Xgap84
33.6 Xtxp348
40.6 Xisp259
42.6 Xtxp100
47.5 Xtxp207
48.4 Xcup26
49.3 Xtxp7
52.9 Xisep841
55.9 Xtxp296
70.5 Xtxp8
77.7 Xisep522
SBI-02
0.1 Undhsbm82
1.4 Xtxp197
2.0 Xcup64
3.9 Xtxp80 Xtxp25
5.0 Xtxp211
5.1 Xtxp50
13.2 Xtxp4
27.8 Xisp346
27.9 Xtxp3 Xtxp72
28.7 Unnsbm154
55.5 Xtxp19
55.9 Xtxp13
57.8 Xisp280
61.4 Xtxp1
61.5 Unnsbm159 Drenhsbm32
62.0 Undhsbm52**
62.7 Xisp336
63.2 Xgap84
67.3 Xtxp348
68.4 Xisp259
69.6 Xtxp100
70.3 Xtxp207 Xtxp7
Xcup26
70.9 Xisep841
71.1 Xtxp296
76.1 Xtxp8
77.8 Xisep522
SBI-03
0.0 Unnhsbm56
3.1 Unnhsbm52**
4.1 Xisp282
10.1 Xcup11
13.2 Xtxp228
16.2 Xtxp116
28.4 Xisp239
38.0 Unnhsbm164
42.8 Xisep101
45.5 Xtxp9
47.9 Xisp277
48.7 Xisp207
51.4 Xisep1012
64.5 Xtxp267#
68.0 Xtxp33
71.7 Xisep114
87.0 Xtxp31
90.2 Xisp307
99.7 Xisp251
100.5 Xisp260
101.5 Stgnhsbm2
102.3 Xtxp218#
104.4 SbAGE01
108.8 Stgnhsbm9
110.0 Xisp355 Xisp332
117.7 Undhsbm314
119.3 Xisp361
122.0 Xtxp38
127.3 Undhsbm209
134.8 Xtxp48
140.2 Xtxp34
141.7 Xcup14
148.5 Drenhsbm10
152.6 Xisp323
SBI-03
0.3 Xisp282
1.5 Undhsbm56
2.0 Xcup11
2.5 Xtxp228 Xtxp116
2.6 Xcup61**
4.6 Xisp239
7.1 Xisep101
7.5 Xisp277
7.6 Xisp207 Xtxp9
8.3 Xisep1012
51.2 Xisep114
51.4 Xtxp33
55.2 Xtxp31
57.8 Xisp307
60.7 Xisp260 Xisp251
Stgnhsbm2
61.4 SbAGE01 Undhsbm209
61.8 Stgnhsbm9
62.9 Xisp355
65.9 Unnsbm164
66.7 Xisp332
68.2 Xisp361 Xtxp38
68.4 Undhsbm314
69.7 Xtxp34
72.5 Xcup14
72.9 Xtxp48
73.8 Drenhsbm10
74.2 Xisp323
SBI-04-1
0.0 Xdhsbm1164
7.9 Xdhsbm1157
10.9 Undhsbm66
11.5 Xisep948
13.6 Xisep224
14.1 Gpsb151
16.1 Xdhsbm1159
17.6 Xdhsbm1173
18.7 Xdhsbm1194
20.6 Xdhsbm1189
21.9 Xdhsbm1172
23.3 Xdhsbm1168
23.9 Xcup48
25.4 Xdhsbm1182
27.5 Xdhsbm1195
36.1 Xcup5
42.1 Xcup23
46.0 Xisp230
48.1 Gpsb50#
49.3 Xtxp26
SBI-04-2
0.0 Xtxp12
16.0 Xtxp343
25.9 Xisep203
30.3 Xdnhsbm43**
42.7 Undhsbm210
52.9 PlcorInt*
61.1 Xisp343#
75.1 Xisep524
77.7 Xtxp51
91.0 Xtxp212
104.4 Xisep234
118.1 Undhsbm265
SBI-04-3
0.0 Xtxp97
1.4 Xtxp104
SBI-04
0.4 Undhsbm66
0.6 Xdhsbm1157
0.7 Xdhsbm1159 Xisep224
0.9 Xdhsbm1164
1.0 Xisep948
1.5 Xcup48
2.0 Xdhsbm1168 Xdhsbm1172
Xdhsbm1173 Xdhsbm1182
Gpsb151
2.1 Xdhsbm1189 Xdhsbm1194
2.2 Xdhsbm1195
3.9 Xcup5
4.3 Xcup23
4.7 Xisp230
4.9 Xtxp26
10.0 Xisep203
48.6 Xtxp12
48.8 Xtxp343
55.7 Undhsbm210
61.2 Xtxp51 Xtxp212
66.1 Xtxp97
66.3 Xtxp104
66.5 Xisep234
66.9 Undhsbm265
67.2 Xisep524

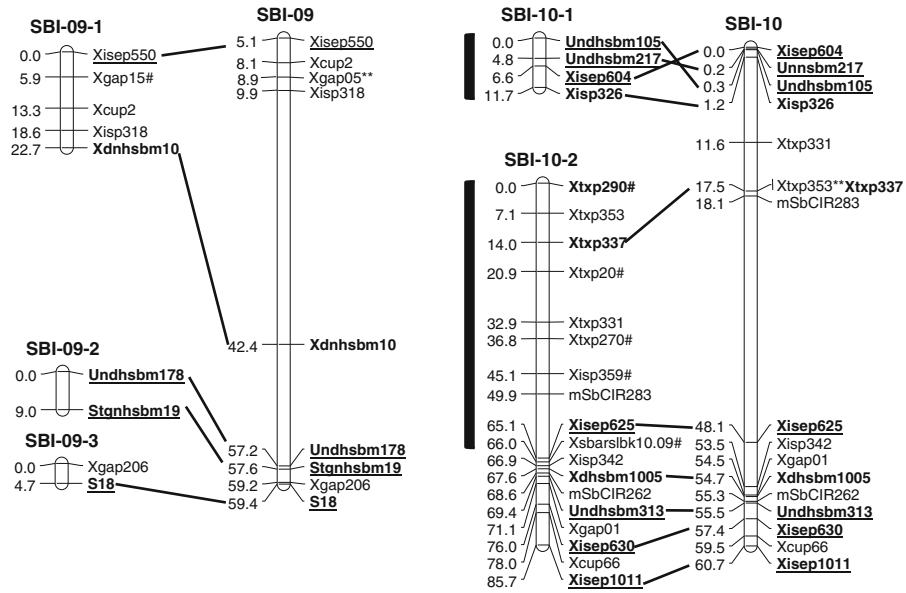
Fig. 1 continued



Putative function of mapped genic-SSRs

BLASTX analysis of 54 mapped genes containing SSRs revealed that 51 (94.4%) showed significant homology with known genes, of which the major proportion (72.9%) code for known proteins and the remaining 13 genes (27.1%) encode uncharacterized proteins (hypothetical) (Table 3). Some of the genes encode important functional proteins involved in drought tolerance mechanisms, including cytochrome P450 71E1, glutamine amidotransferase class-I family protein, ethylene-insensitive-3-like protein, flowering time control protein FCA gamma-like, Avr9 elicitor response protein, heat shock protein 17.2, myb-related protein Hv33, auxin-responsive protein IAA10, Alcohol dehydrogenase 2, farnesylated protein 3, chlorophyll A-B binding protein (CAB), EF-hand Ca^{2+}-binding protein CCD1 and cell elongation

Fig. 1 continued



protein diminuto-related (Rampino et al. 2006; Iuchi et al. 1996; Vierling 1991).

Discussion

SSRs in sorghum unigene database

In the present study, we explored the SSR marker loci in the unigene database of sorghum (NCBI; Pontius et al. 2003). Although the repeat patterns have been studied previously in unigene sequences of sorghum (Parida et al. 2006; Srinivas et al. 2009b), the present study was different from the above studies with respect to stringent length criteria for repeats and a larger number of unigene sequences included. The nature and frequency of SSRs in genomic and EST collections of sorghum has been well documented (Bhattramakki et al. 2000; Jayashree et al. 2006; Srinivas et al. 2008; Ramu et al. 2009). The data obtained in the present study showed the presence of large number of SSRs in the unigene database that can be useful for MAS applications in sorghum breeding. AT-rich motifs are more frequent in mono-, di-, tetra-, penta- and hexamer motifs while CG-rich repeats were found more frequently in trimers. In the case of mononucleotide repeats, for example, the frequency of poly(A) or poly(T) is 52-fold greater than that of poly(G) or poly(C). Similarly, in dimeric repeats, AC, AG and AT are the most abundant compared to CG repeats. Only some of this variation can be explained on the basis of the A/T richness and the relative ease of strand separation compared to C/G tracts (Gur-Arie et al. 2000, Subramanian et al. 2003). The AG repeat, for example, is twice as abundant as the AC repeat. In trimeric repeats, the motifs CCG, AGC, AGG and ACG are the most abundant and the AAT motif is least abundant. The hexamers (64.6%) followed by trimers (20.1%) were the most prevalent class of microsatellites in the unigenes of sorghum. This is similar to earlier observations on the relative abundance of trinucleotide motifs in the EST sequences of cereals (Varshney et al. 2002; Kantety et al. 2002) and other plant genomes (Powell et al. 1996; Gupta et al. 1996) as well as in genomes of higher eukaryotic organisms (Borstnik and Pumpernik 2002; Subramanian et al. 2003), because expansions or deletions in coding regions can be tolerated for tri- and hexa-nucleotide unit repeats, which do not perturb reading frames (Katti et al. 2001). The higher frequency of the trinucleotide repeat motifs than the other classes could be attributed to selection against frame shift mutations which limits expansion of nontriplet microsatellites (Metzgar et al. 2000). The presence of SSRs in genes suggests that they may have a role in gene expression or function. In rice, the length of a poly(CT) SSR in the 5′- untranslated region of the

Table 3 Putative function of genic-SSR markers mapped in the present study

S. No.	Locus	EST accession No./Unigene ID	Linkage group	Sorghum gene	Location	Putative function
1	Undhsbm1	Sbi.63	SBI-01	Sb01g001180	Chromosome_1: 1058885–1061083	Similar to Cytochrome P450 71E1
2	Undhsbm68	Sbi.2622	SBI-01	Sb01g008290	Chromosome_1:7157207–7158750	Similar to Putative uncharacterized protein OJ1124_H03.7
3	Undhsbm90	Sbi.3399	SBI-01	Sb01g011750	Chromosome_1:10629875–10633877	Similar to Glutamine amidotransferase class-I family protein
4	S23	CNL182	SBI-01	Sb01g012740	Chromosome_1:11730337–11733690	Similar to Tubulin beta-2/beta-3 chain
5	Xisep1038	AW565964	SBI-01	Sb01g018800	Chromosome_1:19721291–19725402	Similar to Zinc finger, C3HC4 type family protein
6	Xisep1028	AW747772	SBI-01	Sb01g029930	Chromosome_1:52056562–52057686	Similar to Putative uncharacterized protein
7	Undhsbm24	Sbi.736	SBI-01	Sb01g036740	Chromosome_1:60346941–60350979	Similar to Ethylene-insensitive-3-like protein
8	Undhsbm195	Sbi.8865	SBI-01	Sb01g040640	Chromosome_1:63956959–63957486	Similar to Os08g0239000 protein
9	Undhsbm82	Sbi.3125	SBI-02	Sb02g000230	Chromosome_2:129263–131530	Similar to Tic62 protein, putative, expressed
10	Undhsbm154	Sbi.7032	SBI-02	Sb02g014690	Chromosome_2: 28678566–28691538	Similar to Flowering time control protein FCA gamma-like
11	Undhsbm159	Sbi.7366	SBI-02	Sb02g026360	Chromosome_2: 61465916–61469106	Similar to Putative Avr9 elicitor response protein
12	Xisep841	BE358373	SBI-02	Not in gene	Chromosome_2:70840382–70940481	–
13	Drenhsbm32	CF770135	SBI-02	Sb02g026360	Chromosome_2:61465916–61469106	Similar to Putative Avr9 elicitor response protein
14	Xisep522	BE361646	SBI-02	Sb02g043950	Chromosome_2: 77749505–77754014	Similar to Os07g0693600 protein
15	Xisep101	AW679887	SBI-03	Sb03g006870	Chromosome_3:7097311–7098093	Similar to Heat shock protein 17.2
16	Xisep1012	BE360971	SBI-03	Sb03g007915	Chromosome_3:8300368–8301801	Similar to Putative uncharacterized protein
17	Xisep114	AW745467	SBI-03	Sb03g025455	Chromosome_3:51176407–51176664	Similar to Cysteine-rich extensin-like protein-2
18	Undhsbm56	Sbi.2161	SBI-03	Sb03g001650	Chromosome_3:1479755–1482190	Similar to Putative esterase
19	Undhsbm52	Sbi.2012	SBI-03	Not in gene	Chromosome_2:61933666–62033865	–
20	Stgnhsbm2	CX616697	SBI-03	Sb03g032260	Chromosome_3:60693795–60695579	myb-related protein Hv33, putative, expressed
21	Stgnhsbm9	CD223691	SBI-03	Sb03g033550	Chromosome_3:61820363–61822645	Similar to Putative uncharacterized protein
22	Drenhsbm10	CF772123	SBI-03	Sb03g046730	Chromosome_3: 73788493–73790878	Similar to Probable ubiquitin-fold modifier 1 precursor
23	Undhsbm164	Sbi.7738	SBI-03	Not in gene	Chromosome_3:6589749–6589768	–
24	Undhsbm314	Sbi10880	SBI-03	Sb03g040910	Chromosome_3: 68417828–68418925	Similar to Putative uncharacterized protein
25	Undhsbm209	Sbi.9424	SBI-03	Sb03g041550	Chromosome_3: 69056445–69059782	Similar to Putative uncharacterized protein
26	Xisep224	AW563693	SBI-04	Sb04g000840	Chromosome_4: 698389–701098	Similar to Putative uncharacterized protein
27	Xisep948	BE358472	SBI-04	Sb04g001130	Chromosome_4: 966713–968581	Similar to Catalase isozyme 3
28	Xisep203	BE359080	SBI-04	Sb04g008600	Chromosome_4: 9998074–10011232	Simillar to class III peroxidase
29	Undhsbm66	Sbi.2553	SBI-04	Sb04g000540	Chromosome_4: 372053–375378	Similar to Putative uncharacterized protein OJA1212_C06.19
30	Undhsbm210	Sbi.9457	SBI-04	Sb04g025890	Chromosome_4: 55666981–55674595	Similar to Leucine-rich repeat-like protein
31	Xisep524	BE358788	SBI-04	Sb04g037640	Chromosome_4: 67229617–67232557	Similar to Putative fibrillarin protein

Table 3 continued

S. No.	Locus	EST accession No./Unigene ID	Linkage group	Sorghum gene	Location	Putative function
32	Xisep234	AW680664	SBI-04	Sb04g036660	Chromosome_4: 66478839–66483569	Similar to Putative GTPase activating protein
33	Undhsbm265	Sbi.11534	SBI-04	Sb04g037260	Chromosome_4: 66934638–66936987	Similar to Auxin-responsive protein IAA10
34	Xisep1208	AW287725	SBI-05	Sb05g000940	Chromosome_5: 879883–881239	Hypoxia induced protein conserved region containing protein, expressed
35	Stgnhsbm46	CX611411	SBI-05	Sb05g009350	Chromosome_5:18472314–18475420	Similar to Alcohol dehydrogenase 2
36	Xisep1107	BE125211	SBI-05	Sb05g006520	Chromosome_5: 10622824–10628985	Similar to Expressed protein
37	Xisep443	BE355584	SBI-06	Sb06g026670	Chromosome_6: 55680173 – 55680640	Similar to H0307D04.12 protein
38	Xisep427	AW563837	SBI-06	Sb06g031060	Chromosome_6: 59441882–59445806	Similar to OSJNBa0060D06.16 protein
39	Xisep131	BE357921	SBI-07	Sb07g000380	Chromosome_7: 170550–171704	Similar to Farnesylated protein 3
40	Xisep805	BE352923	SBI-07	Sb07g001680	Chromosome_7: 1933118–1934060	Similar to Putative uncharacterized protein
41	Xisep328	BE359052	SBI-07	Sb07g004290	Chromosome_7: 5440507–5442963	40S ribosomal protein S3
42	Drenhsbm16	CF771342	SBI-08	Sb08g000945	Chromosome_8: 873750–880414	Similar to Putative uncharacterized protein OSJNBa0041F13.17
43	Undhsbm109	Sbi.4081	SBI-08	Sb08g020760	Chromosome_8: 52053371–52059284	Similar to Tetratricopeptide repeat protein, putative, expressed
44	Xisep550	AW746901	SBI-09	Sb09g004270	Chromosome_9: 5060151–5063515	Spliceosomal protein snRNP-U1A/U2B
45	Undhsbm178	Sbi.8301	SBI-09	Sb09g028300	Chromosome_9: 57219361–57222245	Serine/threonine protein kinase
46	Stgnhsbm19	CN144249	SBI-09	Sb09g028720	Chromosome_9: 57560635–57561977	Similar to Chlorophyll A-B binding protein (CAB), putative
47	S18	CNL177	SBI-09	Sb09g030740	Chromosome_9: 59373345–59375990	Similar to Putative uncharacterized protein
48	Undhsbm217	Sbi.9792	SBI-10	Sb10g000390	Chromosome_10: 157912–163241	Similar to Putative uncharacterized protein
49	Xisep604	AW563246	SBI-10	Sb10g000230	Chromosome_10: 31694–32694	Similar to Plastocyanin, chloroplast precursor
50	Undhsbm105	Sbi.4002	SBI-10	Sb10g000470	Chromosome_10: 274080–278793	Similar to Putative uncharacterized protein
51	Xisep625	AW677135	SBI-10	Sb10g021750	Chromosome_10: 48098018–48101431	Similar to Putative uncharacterized protein
52	Undhsbm313	Sbi17757	SBI-10	Sb10g026120	Chromosome_10: 55468789–55472702	Similar to Integral membrane protein-like
53	Xisep630	AW922806	SBI-10	Sb10g027610	Chromosome_10: 57400262–57401004	Similar to EF-hand Ca2 + -binding protein CCD1
54	Xisep1011	AW283079	SBI-10	Sb10g031090	Chromosome_10: 60742768–60746789	Simillar to cell elongation protein diminuto-related

waxy gene is associated with amylose content (Ayres et al. 1997), and in maize SSRs in the 5′-untranslated region of some ribosomal genes have been suggested to be involved in the regulation of fertilization (Dresselhaus et al. 1999).

Genic-SSR polymorphism

Sufficient DNA sequence polymorphisms must be present between the parents in order to construct a dense linkage map. This cannot be overemphasized, for in the absence of DNA polymorphism, segregation analysis and linkage mapping are not possible. The percentage of locus-specific primers (primers amplified with correct product size) was as high at 96.5% in the present study, similar to those in sorghum (Li et al. 2009) and sunflower (Tang et al. 2002). This high percentage can be attributed to the small genome size of sorghum. In the present study, only 19.8% of the unigenic SSRs screened between the parents (M35-1 and B35) out of 302 tested were found to be polymorphic between the two parents, which is slightly less than the polymorphism reported for sorghum EST-SSRs (Ramu et al. 2009) in four mapping population parents (28%). A similar percentage of polymorphism (8–54%) was reported by Thiel et al. (2003) in three different parents of a mapping population in barley.

Genetic linkage map

In the present study, map locations of the 81 new SSRs were determined (Fig. 1). The present map containing 228 SSRs with three morphological markers is the largest map with the greatest number of SSR markers. Earlier, Mace et al. (2009) reported a consensus map of sorghum containing 226 SSRs. The total length of the map was 1235.5 cM, which is comparable with the other linkage maps of sorghum (Bhattramakki et al. 2000; Wu and Haung 2006; Srinivas et al. 2009b; Satish et al. 2009). The positions and order of the SSR loci between this study and that of recent maps of Ramu et al. (2009), Mace et al. (2009), Srinivas et al. (2009b) and Bhattramakki et al. (2000) are in good agreement. The bridge genomic SSR loci were distributed over all 21 linkage groups identified in this study and facilitated alignment of novel SSRs to the chromosomes and subsequent identification of their positions on the linkage map. Except for the two chromosomes (SBI-03 and SBI-05) which had one linkage group, five other chromosomes (SBI-01, SBI-06, SBI-07, SBI-08 and SBI-10) had two fragments each and the other three (SBI-02, SBI-04 and SBI-09) had three linkage group fragments. More markers are required to combine these fragmented linkage groups. Few rearrangements in consecutive markers were observed in the present study. This is not surprising as these changes are possible in the construction of linkage maps involving closely linked markers. In general, loci order should be same, while the genetic distances can vary among different maps of same species.

A difference ratio was calculated to compare the genetic distance between the present genetic map with the recent maps of Srinivas et al. (2009b) and Wu and Haung (2006), and the consensus map of Mace et al. (2009). The overall difference ratios in genetic distance between the present map with other maps are comparable to ratios calculated by Wu and Haung (2006) and Mace et al. (2009). The low difference ratios observed indicate that there is good agreement in overall distance between common marker pairs. It was observed that map estimates are less important than marker order, as map distances do vary between different genetic linkage maps by several centimorgans (Van et al. 2005), and that the marker order is the most critical feature for further applications of the map, including map-based cloning and synteny-based marker development. The difference ratios also vary for each chromosome across maps, with SBI-08 having the lowest difference ratios and SBI-03 having the highest (Supplementary Table 3). The difference ratio observed for some chromosomes, specifically for SBI-01, SBI-03, SBI-04 and SBI-06 of the present map with the map of Srinivas et al. (2009b), SBI-04 and SBI-06 in comparison with the map of Mace et al. (2009), and SBI-04 and SBI-07 with the map of Wu and Haung (2006), was high. This can be attributed to the low number of sampling intervals available for comparison per linkage group. For example, linkage groups SBI-06 and SBI-07 in the map of Srinivas et al. (2009b) had a single common interval with present map. In some cases, there were no intervals in common for particular chromosomes across maps; for example, linkage groups SBI-05 and SBI-09 of

Srinivas et al. (2009b) had no shared intervals with the current map.

In the present study, 60 markers showed significant segregation distortion ($P < 0.1$) from the expected Mendelian ratio (1:1), of which 30 markers clustered on three different chromosomes during linkage analysis, one cluster each on SBI-01, SBI-04 and two clusters on SBI-10. In all three chromosomes, distorted alleles were from the female parent M35-1. Of the three chromosomes hosting significant segregation distortion identified in the present study, one chromosome (SBI-01) was previously detected to harbor significantly distorted loci (Menz et al. 2002; Srinivas et al. 2008). Strong and consistent segregation distortion in one region cannot be explained by sampling errors but rather suggests selection favoring one parental allele. Several genetic factors related to compatibility, gamete or zygote viability, deleterious recessive alleles, self-incompatibility alleles, structural rearrangements and differences in DNA content (Moretzsohn et al. 2005; Taylor and Ingvarsson 2003) and/or strong effects on general fitness could reside in this region. Menz et al. (2002) suggested that segregation distortion in chromosome SBI-01 could be due to the presence of shattering genes located in this chromosomal region. Menz et al. (2002) also reported distorted segregating loci on chromosome SBI-10. The segregation distortion of markers has also been described in many DNA marker-based linkage maps, including sorghum (Chittenden et al. 1994; Dufour et al. 1997; Wu and Haung 2006). In the linkage analysis, the deviating SSRs were retained since their map positions in the present map were consistent with the earlier reports (Bhattramakki et al. 2000, Haussmann et al. 2002a, b, Wu and Haung 2006), their BLAST position with sorghum genome sequence, and they co-segregated very tightly with the known markers. The majority of newly mapped loci in this study are evenly distributed throughout the linkage map, covering the gaps of the previous maps (Srinivas et al. 2009a, b; Mace et al. 2009; Bhattramakki et al. 2000), especially on the upper portion of SBI-01 near Xtxp302, on SBI-04 around Xcup48, on SBI-05 around Gpsb17, on SBI-09 around Xgap206, and on SBI-10 around Xtxp331 and Xgap1 (Fig. 1). For example, 14 markers were mapped to the marker-sparse region on SBI-04-1 and extended the ends of this linkage group from 23.1 cM (Srinivas et al. 2009a, b) to 49.3 cM. Similarly, the mapping of markers Undhsbm82 on SBI-02-1 and Xtxp127 on SBI-06 also extended the length of the linkage groups. The genetic map of the present study, with 81 additional markers and 54 genic markers, will have applications in functional diversity studies, association mapping studies and QTL mapping studies for staygreen and other agronomic traits in sorghum.

Molecular markers can be mapped by using either classical genetic mapping, which is based on segregating marker data of a population, or in-silico mapping. Classical genetic mapping is the most reliable way to map molecular markers, but it is costly and time-consuming to develop populations and to genotype individuals. Sometimes, a few markers cannot be mapped because of lack of polymorphism between the parents of the mapping population. In contrast, the in-silico mapping strategy is convenient, and saves labour and cost. However, it is restricted by the number of genome sequences, such as ESTs, sequenced bacterial artificial chromosomes, or at least by primer sequences. In-silico mapping of the 209 genetically mapped SSR markers on the aligned sorghum genome sequence (Paterson et al. 2009) was performed in the present study using their primer sequence information and marker positions matched with the conventional linkage map, both in terms of chromosome arm location and order. The order of these SSRs from the in-silico mapping agrees with the linkage map for all ten sorghum linkage groups. Recently, 5012 (Yonemaru et al. 2009) and 1758 (Li et al. 2009) genomic SSRs were mapped in silico onto the aligned sorghum genome sequence assembly. In the present study, 18 SSR markers (8.0%) out of 227, though mapped by linkage analysis onto the correct linkage groups, produced no hits with the genome sequence in in-silico analysis. Only five (2.2%) SSR markers of the present genetic map were mapped to different chromosomal regions by in-silico mapping. Ramu et al. (2009) reported that 12.5% of mapped SSR markers were mapped to linkage groups other than those expected by in-silico mapping. The difference in the mapping results of in-silico and classical genetic mapping may be caused by either misassembly of genome sequences or by weak linkage in classical genetic mapping. However, such differences, due to ancestral genome duplication events or transposition within the genome at more than one location, cannot be ignored. The in-silico-predicted physical map positions of markers could help to

saturate the marker-sparse regions of any chromosome. Thus, the availability of genome sequences and other marker information could help in selecting or developing markers for targeted mapping, not only in a given crop but also in related crops for which sufficient genomic tools are not yet available in the public domain.

Mapping of genes with known functions on the linkage map permits evaluation of associations between genic markers and QTL of any trait (Aubert et al. 2006) and provides a basis both for understanding the genetic basis of a trait (Matthews et al. 2001; Zhang et al. 2004) and for MAS in breeding programs (Srinivas et al. 2009a, b). In this study, 54 genic markers were mapped to different chromosomes. A unigene marker Undhsbm24 coding for ethylene-insensitive-3-1 (EIL-1) protein, a key transcription regulator of ethylene biosynthesis (Chen et al. 2009; Solano et al. 1998), was mapped on linkage group SBI-01. Similarly, other functional proteins that are involved in stress-related metabolism were mapped on the linkage map. For instances, EST-derived SSR markers Stgnhsbm19 (mapped on SBI-09), Stgnhsbm46 (mapped on SBI-05) and Xisep101 (mapped on SBI-03) were identified to code for chlorophyll A-B binding protein (CAB), alcohol dehydrogenase 2 (*adh2*) and heat shock protein 17.2, respectively, which are involved in drought stress-related mechanisms (Rampino et al. 2006; Iuchi et al. 1996; Vierling 1991). In this respect, the mapped genic markers in this study not only provide reproducible, co-dominant and easily scorable markers, but also hypothetically include candidate genes that have the potential of being casually linked to traits of agronomic importance. Therefore, the 54 genic SSRs mapped on the linkage map in this study may assist the study of the association of the molecular variability of the genes with the phenotypic variability, if any, of traits of agronomic importance in sorghum.

Map positions of morphological traits

In the present study, the genes controlling three major genes, pericarp colour (*Pericarp)*, plant colour intensity (*PlcorInt)* and yellow midrib colour (*Ymrco*), were mapped on chromosomes SBI-01, SBI-04 and SBI-06, respectively. The map position of the *Pericarp* locus on SBI-01 was similar to that reported by Knoll et al. (2007) on the same chromosome but at a different position, while the map positions of the two other morphological markers (*PlcorInt* and *Ymrco*) were reported for the first time in sorghum. These markers are under monogenic control and are independent of environment, and therefore they can be effectively used in MAS if associated with any of the QTL and will add to the list of major genes mapped in sorghum (Mace and Jordan 2010).

In conclusion, the present study reports the identification, development and mapping of new microsatellite markers from the unigene database of sorghum. The genic-SSRs developed in the study will have applications in functional diversity studies, association mapping studies and QTL mapping studies for drought and other traits of economic importance in sorghum.

Acknowledgments The authors gratefully acknowledge the Department of Biotechnology (DBT), Government of India (GOI), for supporting this work under the network projects "Development and applications of biotechnological tools for millet improvement" under grant no. BT/PR 6031/AGR/02/307/2005, and the Director, Directorate of Sorghum Research (DSR), Rajendranagar, Hyderabad for the facilities to undertake the study.

References

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–559

Aubert G, Morin J, Jacquin F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune-Henaut I, Huguet T, Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theor Appl Genet 112: 1024–1041

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94:773–781

Berhan AM, Hulbert SH, Bulter LG, Bennetzen JL (1993) Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*. Theor Appl Genet 86:598–604

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

Binelli G, Gianfranceschi L, Po ME, Taramino G, Busso C, Stenhouse J, Ottaviano E (1992) Similarity of maize and sorghum genomes as revealed by maize RFLP probes. Theor Appl Genet 84:10–16

Boivin K, Deu M, Rami JF, Trouche G, Hamon P (1999) Towards a saturated sorghum map using RFLP and AFLP markers. Theor Appl Genet 98:320–328

Borstnik B, Pumpernik D (2002) Tandem repeats in protein coding regions of primate genes. Genome Res 12:909–915

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, Lin YR, Liu SC, Luo L, Marler BS, Ming R, Mitchell SE, Qiang D, Reischmann K, Schulze SR, Skinner DN, Wang YW, Kresovich S, Schertz KF, Paterson AH (2003) A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzales-Candelas F, Kresovitch S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 93:190–198

Calviño M, Miclaus M, Bruggmann R, Messing J (2009) Molecular markers for sweet sorghum based on micro-array expression data. Rice, doi 10.1007/s12284-009-9029-8

Chen H, Xue L, Chintamanani S, Germain H, Lin H, Cui H, Cai R, Zuo J, Tang X, Li X, Guo H, Zhou JM (2009) ETHYLENE INSENSITIVE3 and ETHYLENE INSENSITIVE3-LIKE1 repress SALICYLIC ACID INDUCTION DEFICIENT2 expression to negatively regulate plant innate immunity in Arabidopsis. Plant Cell 21:2527–2540

Childs KL, Klein RR, Klein PE, Morishige DT, Mullet JE (2001) Mapping genes on an integrated sorghum genetic and physical map using cDNA selection technology. Plant J 27:243–255

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed RFLP map of *Sorghum bicolor* × *S. propinquum*, suitable for high density mapping, suggests ancestral duplication of sorghum chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

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

Crasta OR, Xu WW, Rosenow DT, Mullet J, Nguyen HT (1999) Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol Gen Genet 262:579–588

Cummings CJ, Zoghbi HY (2000) Trinucleotide repeats: mechanisms and pathophysiology. Annu Rev Genom Hum Genet 1:281–328

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19:4008

Dresselhaus T, Cordts S, Heuer S, Sauter M, Lörz H, Kranz E (1999) Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles. Mol Gen Genet 261:416–427

Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, Lanaud C, Glaszmann JC, Hamon P (1997) Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor Appl Genet 94:409–418

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99:125–132

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422

FAOSTAT (2009) http://faostat.fao.org/

Fernandez MGS, Hamblin MT, Li L, Rooney WL, Tunistra MR, Kresovich S (2008) Quantitative trait loci analysis of endosperm color and carotenoid content in sorghum grain. Crop Sci 48:1732–2743

Fritz AK, Caldwell S, Worral WD (1999) Molecular mapping of Russian wheat aphid resistance from triticale accession, PI 386156. Crop Sci 39:1707-1710

Gong L, Deng Z (2010) EST-SSR markers for gerbera (*Gerbera hybrida*). Mol Breed 26:125–132. doi:10.1007/s11032-009-9380-x

Goulden CH (1939) Problems in plant selection. In: Burnett RC (ed) Proceedings international genetics congress, 7th, Edinburgh, 1939. Cambridge University Press, Edinburgh, pp 132–133

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics 27:857–874

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

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 70:45–54

Gupta HS, Agrawal PK, Mahajan V, Bisht GS, Kumar A, Verma P, Srivastava A, Saha S, Babu R, Pant MC, Mani VP (2009) Quality protein maize for nutritional security: rapid development of short duration hybrids through molecular marker assisted breeding. Curr Sci 96:230–237

Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM, Kashi Y (2000) Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. Genome Res 10:62–71

Harris K, Subudhi PK, Borrell A, Jordan DR, Rosenow DT, Nguyen HT, Klein PE, Klein RR, Mullet J (2007) Sorghum staygreen QTL individually reduce post-flowering drought induced leaf senescence. J Exp Bot 58:327–338

Haussmann BIG, Hess DE, Seetharama N, Welz HG, Geiger HH (2002a) Construction of a combined sorghum linkage map from two recombinant inbred population using AFLP, SSR, RFLP, and RAPD markers, and comparison with other sorghum maps. Theor Appl Genet 105:629–637

Haussmann BIG, Mahalakshmi V, Reddy BVS, Seetharama N, Hash CT, Geiger HH (2002b) QTL mapping of staygreen in two sorghum recombinant inbred populations. Theor Appl Genet 106:133–142

Heesacker A, Kishore VK, Gao WX, Tang SX, Kolkman JM, Gingle A, Matvienko M, Kozik A, Michelmore RM, Lai Z, Rieseberg LH, Knapp SJ (2008) SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. Theor Appl Genet 117:1021–1029

Hulbert SH, Richter TE, Axtell JD, Bennetzen JL (1990) Genetic mapping and characterization of sorghum and

related crops by means of maize DNA probes. Proc Natl Acad Sci USA 87:4251–4255

Iuchi S, Yamaguchi-Shinozaki K, Urao T, Terao T, Shinozaki K (1996) Novel drought-inducible genes in the highly drought-tolerant cowpea: cloning of cDNAs and analysis of the expression of the corresponding genes. Plant Cell Physiol 37:1073–1082

Jayashree B, Punna R, Prasad P, Bantte K, Hash CT, Chandra S, Hoisington DA, Varshney RK (2006) A database of simple sequence repeats from cereal and legume expressed sequence tags mined *in silico*: survey and evaluation. In Silico Biol 6:0054. http://www.bioinfo.de/isb/2006/06/0054/

Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 18:1161–1167

Kim JS, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169:1169–1173

Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MA, Franks CD, Klein PE (2008) The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Sci 48:S12–S26

Knoll J, Gunaratna N, Ejeta G (2007) QTL analysis of early-season cold tolerance in sorghum. Theor Appl Genet 116:577–587, doi 10.1007/s00122-007-0692-0

Kong L, Dong J, Hart GE (2000) Characteristics, linkage map positions, and allelic differentiation of Sorghum bicolor (L.) Moench. DNA simple-sequence repeats (SSRs). Theor Appl Genet 101:438–448

Li M, Yuyama N, Luo L, Hirata M, Cai H (2009) *In silico* mapping of 1758 new SSR markers developed from public genomic sequences for sorghum. Mol Breed 24:41–47, doi:10.1007/s11032-009-9270-2

Lopez C, Jorge V, Pie′gu B, Mba C, Cortes D, Restrepo S, Soto M, Laudie′ M, Berger C, Cooke R, Delseny M, Tohme J, Verdier V (2004) A unigene catalogue of 5700 expressed genes in cassava. Plant Mol Biol 56:541–554

Mace ES, Jordan DR (2010) Location of major effect genes in sorghum (*Sorghum bicolor* (L.) Moench). Theor Appl Genet 121:1339–1356, doi:10.1007/s00122-010-1392-8

Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A (2008) DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC Genom 9:26–30

Mace ES, Rami JF, Bouchet S, Klein PE, Klein RR, Kilian A, Wenz P, Xia L, Halloran K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biol 9:13, doi: 10.1186/1471-2229-9-13

Matthews BF, Devine TE, Weisemann JM, Beard HS, Lewers KS, McDonald MH, Park YB, Maiti R, Lin JJ, Kuo J, Pedroni MJ, Cregan PB, Saunders JA (2001) Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. Crop Sci 41:516–521

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9(6):199–207

Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of Sorghum bicolor (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Mol Biol 48:483–499

Metzgar D, Bytof J, Wills C (2000) Selection against frame shift mutations limits microsatellite expansion in coding DNA. Genome Res 10:72–80

Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SC, Gimenes MA, Martins WS, Valls JF, Grattapaglia D, Bertioli DJ (2005) A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (*Fabaceae*). Theor Appl Genet 111:1060–1071

Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, Kresovich S (2008) Genetic improvement of sorghum as a biofuel feedstock I: QTL for stem sugar and grain nonstructural carbohydrates. Crop Sci 48:2165–2179

Ottoman RJ, Hane DC, Brown CR, Yilma S, James SR, Mosley AR, Crosslin JM, Vales MI (2009) Validation and implementation of marker-assisted selection (MAS) for PVY resistance (Rygene) in a tetraploid potato breeding program. Am J Potato Res 86:304–314

Parh DK, Jordan DR, Aitken EAB, Mace ES, Jun-ai P, McIntyre CL, Godwin ID (2008) QTL analysis of ergot resistance in sorghum. Theor Appl Genet 117:369–382

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

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman WareD, Westho VP, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Pontius JU, Wagner L, Schuler GD (2003) UniGene: a unified view of the transcriptome. In: McEntyre J, Ostell J (eds) The NCBI Handbook. National Library of Medicine (US), NCBI, Bethesda, MD

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Pratt LH, Liang C, Shah M, Sun F, Wang H, Reid SP, Alan RG, Paterson AH, Wing R, Dean R, Klein R, Nguyen HT, Hi Ma, Zhao X, Daryl TM, Mullet JE, Cordonnier-Pratt M (2005) Sorghum expressed sequence tags identify signature genes for drought, pathogenesis, and skotomorphogenesis

from a milestone set of 16, 801 unique transcripts. Plant Physiol 139:869–884
Rampino P, Spano G, Pataleo S, Mita G, Napier JA, Fonzo ND, Shewry PR, Perrotta C (2006) Molecular analysis of a durum wheat 'staygreen' mutant: expression pattern of photosynthesis-related genes. J Cereal Sci 43:160–168
Ramu P, Kassahun B, Senthilvel S, Ashok Kumar C, Jayashree B, Folkertsma RT, Ananda Reddy L, Kuruvinashetti MS, Haussmann BIG, Hash CT (2009) Exploiting rice-sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. Theor Appl Genet 119:1193–1204
Ritter KB, Jordan DR, Chapman SC, Godwin ID, Mace ES, McIntyre CL (2008) Identification of QTL for sugar-related traits in a sweet × grain sorghum *(Sorghum bicolor* L. Moench) recombinant inbred population. Mol Breed 22:367–384
Saghai-Maroof 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 81:8014–8018
Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT (2002) Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* L. Moench). Plant Mol Biol 48:713–726
Satish K, Srinivas G, Madhusudhana R, Padmaja PG, Nagaraja Reddy R, Murali Mohan S, Seetharama N (2009) Identification of quantitative trait loci for resistance to shoot fly in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 119(8):1425–1439, doi 10.1007/s00122-009-1145-8
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
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
Solano R, Stepanova A, Chao Q, Ecker JR (1998) Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESP ONSE—FACTOR1. Genes Dev 12:3703–3714
Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339–1348
Srinivas G, Satish K, Murali Mohan S, Nagaraja Reddy R, Madhusudhana R, Balakrishna D, Venkatesh Bhat B, Howarth CJ, Seetharama N (2008) Development of genic-microsatellite markers for sorghum staygreen QTL using a comparative genomic approach with rice. Theor Appl Genet 117:283–296
Srinivas G, Satish K, Madhusudhana R, Seetharama N (2009a) Exploration and mapping of microsatellite markers from subtracted drought stress ESTs in *Sorghum bicolor* (L.) Moench. Theor Appl Genet 118:703–717
Srinivas G, Satish K, Madhusudhana R, Nagaraja Reddy R, Murali Mohan S, Seetharama N (2009b) Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum. Theor Appl Genet 118:1439–1454
Stam P, Van Ooijen JW (1995) JoinMapTM version 2.0: software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen
Subramanian S, Madgula VM, George R, Mishra RK, Pandit MW, Kumar CS, Singh L (2003) Triplet repeats in human genome: distribution and their association with genes and other genomic regions. Bioinformatics 19:549–552
Subudhi PK, Rosenow DT, Nguyen HT (2000) Quantitative trait loci for the stay green trait in sorghum (*Sorghum bicolor* L. Moench): consistency across genetic backgrounds and environments. Theor Appl Genet 101:733–741
Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1130
Taramino G, Tarchini R, Ferrario S, Lee M, Pe ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95:66–72
Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. Genetica 117:27–35
Temnykh S, Park WD, Ayers N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712
Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422
Van Ooijen JW, Voorrips RE (2001) JoinMap 3.0. Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands
Van OSH, Stam P, Visser RGF, Van Eck HG (2005) RECORD: a novel method for ordering loci on a genetic linkage map. Theor Appl Genet 112:30–40
Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546
Vierling E (1991) The roles of heat shock proteins in plants. Annu Rev Plant Physiol Plant Mol Biol 42:579–620
Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93(1):77–78
Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132:1119–1130
William HM, Trethowan R, Crosby-Galvan EM (2007) Wheat breeding assisted by markers: CIMMYT's. experience. Euphytica 157:307–319
Wu YQ, Haung Y (2006) An SSR genetic map of *Sorghum bicolor* (L) Monch and its comparison to a published genetic map. Genome 50:84–89
Xu GW, Magill CW, Schertz KF, Hart GE (1994) A RFLP linkage map of *Sorghum bicolor* (L.) Moench. Theor Appl Genet 89:139–145
Xu W, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen NT (2000) Molecular mapping of QTLs conferring staygreen in grain sorghum (*Sorghum bicolor* L. Moench). Genome 43:461–469

Yonemaru J-I, 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, doi:10.1093/dnares/dsp005

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor Appl Genet 108:1131–1139