Theor Appl Genet (2009) 119:1193–1204
DOI 10.1007/s00122-009-1120-4

ORIGINAL PAPER

Exploiting rice–sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map

P. Ramu · B. Kassahun · S. Senthilvel · C. Ashok Kumar · B. Jayashree · R. T. Folkertsma · L. Ananda Reddy · M. S. Kuruvinashetti · B. I. G. Haussmann · C. T. Hash

Received: 6 March 2009 / Accepted: 20 July 2009 / Published online: 8 August 2009

Abstract The sequencing and detailed comparative functional analysis of genomes of a number of select botanical models open new doors into comparative genomics among the angiosperms, with potential benefits for improvement of many orphan crops that feed large populations. In this study, a set of simple sequence repeat (SSR) markers was developed by mining the expressed sequence tag (EST) database of sorghum. Among the SSR-containing sequences, only those sharing considerable homology with rice genomic sequences across the lengths of the 12 rice chromosomes were selected. Thus, 600 SSR-containing sorghum EST sequences (50 homologous sequences on each of the 12 rice chromosomes) were selected, with the intention of providing coverage for corresponding homologous regions of the sorghum genome. Primer pairs were designed and polymorphism detection ability was assessed using parental pairs of two existing sorghum mapping populations. About 28% of these new markers detected polymorphism in this 4-entry panel. A subset of 55 polymorphic EST-derived SSR markers were mapped onto the existing skeleton map of a recombinant inbred population derived from cross N13 × E 36-1, which is segregating for *Striga* resistance and the stay-green component of terminal drought tolerance. These new EST-derived SSR markers mapped across all 10 sorghum linkage groups, mostly to regions expected based on prior knowledge of rice–sorghum synteny. The ESTs from which these markers were derived were then mapped *in silico* onto the aligned sorghum genome sequence, and 88% of the best hits corresponded to linkage-based positions. This study

Communicated by T. Sasaki.

P. Ramu and B. Kassahun contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1120-4) contains supplementary material, which is available to authorized users.

P. Ramu · B. Kassahun · S. Senthilvel · C. Ashok Kumar · B. Jayashree · R. T. Folkertsma · C. T. Hash (✉)
M. S. Swaminathan Applied Genomics Laboratory,
International Crops Research Institute
for the Semi-Arid Tropics (ICRISAT),
ICRISAT-Patancheru PO, Hyderabad,
Andhra Pradesh 502 324, India
e-mail: c.hash@cgiar.org

P. Ramu · L. A. Reddy
Department of Genetics, Osmania University,
Hyderabad, Andhra Pradesh 500 007, India

B. Kassahun · M. S. Kuruvinashetti
University of Agricultural Sciences,
Dharwad, Karnataka 580 005, India

B. Kassahun
College of Agriculture and Veterinary Medicine,
Jimma University, P.O. Box 307, Jimma, Ethiopia

Present Address:
C. Ashok Kumar
Department of Plant Pathology & Crop Physiology,
Louisiana State University, Baton Rouge, LA 70803, USA

R. T. Folkertsma
Department of Phytopathology, De Ruiter Seeds,
Leeuwenhoekweg 52, 2661 CZ, Bergschenhoek, The Netherlands

B. I. G. Haussmann
International Crops Research Institute for the Semi-Arid Tropics
(ICRISAT), BP 12404, Niamey, Niger

demonstrates the utility of comparative genomic information in targeted development of markers to fill gaps in linkage maps of related crop species for which sufficient genomic tools are not available.

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench, $2n = 20$] is one of the most important cereal crops and well adapted to harsh environments characterized by drought and high temperatures. In addition to being an important source of calories for millions of poor people living in the semi-arid tropics of Africa and Asia, and an essential livestock feed there and elsewhere, sorghum has been identified as a potential source of bio-fuel (Antonopoulou et al. 2008). Sorghum is a model organism for tropical grasses having the 'C_4' photosynthetic pathway and is a logical complement to the 'C_3' grass *Oryza sativa* (Kresovich et al. 2005), which was the first monocot with a completely sequenced genome (Goff et al. 2002; Yu et al. 2002; International Rice Genome Sequencing Project 2005). Sorghum has a relatively small genome (~760 Mb; Arumuganathan and Earle 1991) that is less complex than those of other major C_4 crops like maize (*Zea mays* L.) and sugarcane (*Saccharum officinarum*). Genome sequencing of sorghum was initiated at the end of 2005 through the 'Community sequencing program of Department of Energy-Joint Genome Institute, USA' and during the course of this study the complete aligned genome sequence (8×) information for this species was made available to public use (Paterson et al. 2009; http://www.jgi.doe.gov/sorghum). This has provided an opportunity to better understand genome organization in sorghum, its wild relatives including Johnson grass (*Sorghum halepense*), and even maize.

Sorghum is the first species in the crop improvement mandate of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) for which a reasonable amount of genomic tools are available that can be exploited for applied crop improvement. Construction of molecular marker-based linkage maps in sorghum started during the 1990s with restriction fragment length polymorphism (RFLP) markers (Hulbert et al. 1990; Pereira et al. 1994; Chittenden et al. 1994; Ragab et al. 1994; Xu et al. 1994; Dufour et al. 1997; Tao et al. 1998a; Peng et al. 1999) followed by maps adding other marker systems like amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), and recently diversity array technology (DArT) markers (Boivin et al. 1999; Kong et al. 2000; Bhattramakki et al. 2000; Klein et al. 2000; Subudhi and Nguyen 2000; Haussmann et al. 2002a; Menz et al. 2002; Bowers et al. 2003; Mace et al. 2008). Several genomic regions associated with important agronomic traits including a range of abiotic stress tolerances [primarily mid-season and terminal drought tolerance (Tuinstra et al. 1996, 1997; Crasta et al. 1999; Subudhi et al. 2000; Tao et al. 2000; Xu et al. 2000; Kebede et al. 2001; Haussmann et al. 2002b; Sanchez et al. 2002; Harris et al. 2007) and aluminum tolerance (Magalhaes et al. 2004)], host plant resistances to hemi-parasitic weeds [*Striga* spp. (Haussmann et al. 2004)], insect pests [aphids (Agrama et al. 2002; Katsar et al. 2002; Nagaraj et al. 2005; Wu et al. 2007), midge (Tao et al. 2003), and shoot fly (Folkertsma et al. 2003)], and diseases [of both foliage (Oh et al. 1996; Tao et al. 1998b; Nagy et al. 2007) and panicles (Klein et al. 2001)], and a range of traits related to grain quality (Lijavetsky et al. 2000), crop phenology [height and maturity (Lin et al. 1995; Pereira and Lee 1995; Klein et al. 2001; Feltus et al. 2006a)], and yield components (Hart et al. 2002; Hicks et al. 2002) have been mapped. Now high-density genetic maps are available for sorghum (Klein et al. 2000; Bowers et al. 2003). However, these maps have some significant gaps and many genomic regions are covered by only AFLP and RFLP markers. At the onset of this study there were circa 300 publicly available primer pair sequences for mapped sorghum SSR loci (Brown et al. 1996; Taramino et al. 1997; Kong et al. 2000; Bhattramakki et al. 2000; Menz et al. 2002; Schloss et al. 2002).

SSR markers have clear advantages over RFLP and AFLP markers in terms of technical simplicity, throughput level and automation (Varshney et al. 2005). SSRs are the preferred marker system for many breeding applications. Hence, enriching the existing sorghum linkage map with more SSR markers is a valuable objective for the sorghum breeding community globally. Conventional SSR marker development is a costly and time-consuming process. Thanks to the availability of genomic or expressed sequence tag (EST) sequences in public databases and the recent advent of bioinformatics tools, SSR marker development has become easier and more cost-effective (e.g., Jayashree et al. 2006). In the past, SSR markers have been successfully developed by mining EST databases in several crops (reviewed in Varshney et al. 2005). EST-SSRs are a highly valued marker system as they are developed from transcribed portions of the genome and often target functional diversity. They are also superior in terms of cross-species transferability, and thus are well suited for application in phylogenetic analysis and comparative genome mapping. The transfer rate of EST-SSR markers from sorghum to paspalum (*Paspalum* spp.) and to maize was 68 and 61%, respectively (Wang et al. 2005).

The rice genome exhibits substantial collinearity with the genomes of other grasses, such as sorghum, maize, wheat (*Triticum aestivum* L.), and barley (*Hordeum vulgare* L.) (Ahn et al. 1993; Chen et al. 2002; Devos and Gale 1997; Tarchini et al. 2000). A direct comparison of the

genetic linkage maps of sorghum and rice has been performed by Ventelon et al. (2001) using the mapping information of a common set of RFLP probes. Klein et al. (2003) indicated that the overall architecture of sorghum chromosome 3 and rice chromosome 1 has remained largely intact with the exception of one major rearrangement. The sorghum genome has now been reasonably aligned with that of rice through comparative genetic, cytogenetic, and physical mapping approaches (Draye et al. 2001; Bowers et al. 2005, Kim et al. 2005a, Paterson et al. 2009). In this study, we attempted to develop a set of EST-SSR markers that are expected to provide coverage for gene-rich regions of the entire nuclear genome using comparative genomic information from rice as the sorghum genome sequence information was not available when the study was initiated.

Materials and methods

Data mining of sorghum ESTs

Sorghum EST sequences were downloaded in FASTA format from the J. Craig Venter Institute [formerly, The Institute for Genome Research (TIGR)] database (http://www.tigr.org) as of May 2004. The identification of microsatellites from these EST sequences was carried out using SSRIT (www.gramene.org/db/searches/ssrtool), and simple scripts (written in Visual Basic) to parse SSRIT output into the relational database. CAP3 (Huang and Madan 1999) was used to identify the non-redundant EST sequences. Non-redundant sorghum EST sequences containing SSRs were then obtained from this local database and used in this study.

Selection of candidate ESTs and primer design

Non-redundant sorghum EST sequences containing microsatellites were BLAST searched against the rice genome sequence available in the GRAMENE database (http://www.gramene.org/db/searches/blast) as of May 2004. Each rice chromosome was divided into five arbitrary regions viz., 'top', 'middle, 'bottom', 'between top and middle', and 'between middle and bottom'. The ten top-scoring sorghum EST sequences for each region of the rice chromosome were picked. Thus, a total of 600 SSR-containing EST sequences of sorghum (10 each for each region of the 12 rice chromosomes) having sequence homology with rice sequences were identified. These 600 candidate EST-SSR sequences were expected to sparsely cover the entire nuclear genome of rice, and thereby the corresponding sorghum genomic regions. Primer pairs were designed from the selected 600 non-redundant sorghum EST-SSR sequences using Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) after masking the repeat units. The primer pairs were designed with the following conditions: primer length (min-18nt, opt-22nt, max-24nt), Tm (min-54°C, opt-57°C, max-60°C), and GC content (min-45, opt-50, max-60%).

Primer optimization and polymorphism assessment

PCR conditions for the EST-SSR primer pairs were optimized using template DNA of genetically diverse parental pairs of two sorghum mapping populations available at ICRISAT, Patancheru viz., N13 × E 36-1 and BTx623 × IS 18551. N13 is *Striga* resistant while E 36-1 is stay-green, drought tolerant, and *Striga* susceptible. BTx623 is an elite hybrid parental line chosen for genome sequencing (Paterson et al. 2009). It is susceptible to shoot fly and being used as the recurrent parent for ICRISAT's exploratory shoot fly resistance marker-assisted backcrossing program. IS 18551 is a shoot fly resistant donor parent being used by ICRISAT for shoot fly resistance QTL mapping and marker-assisted backcrossing. DNA of parental lines was isolated using a high-throughput DNA extraction protocol as reported by Mace et al. (2003) and normalized to a working concentration of 5 ng/µl.

PCR was performed in 5 µl reaction volume with final concentrations of 2.5 ng DNA, 2 mM $MgCl_2$, 0.1 mM of dNTPs, 1× PCR buffer, 0.4 pM of each primer, and 0.1 U of *Taq* DNA polymerase (AmpliTaq Gold®, Applied Biosystems, USA) in a GeneAmp® PCR System 9700 thermal cycler (Applied Biosystems, USA) with the following cyclic conditions: initial denaturation at 94°C for 15 min (to activate *Taq* DNA polymerase) then 10 cycles of denaturation at 94°C for 15 s, annealing at 61°C for 20 s (temperature reduced by 1°C for each cycle), and extension at 72°C for 30 s. This was followed by 34 cycles of denaturation at 94°C for 10 s, annealing at 54°C for 20 s, and extension at 72°C for 30 s with the final extension of 20 min at 72°C. Amplified PCR products were resolved on 6% native polyacrylamide gels coupled with silver staining as described by Tegelstrom (1992).

Mapping of EST-SSRs

The polymorphic markers were subsequently mapped using recombinant inbred lines (RILs) of N13 × E 36-1 (Haussmann et al. 2004). The markers polymorphic between N13 and E 36-1 were surveyed on a subset of 94 F_3-derived F_5 RILs. The segregation data were used to place the new markers on to the existing skeleton map with 164 previously mapped RFLP, AFLP, RAPD, and SSR markers (Haussmann et al. 2002a, 2004) at LOD score of 3.0 using Mapmaker 3.0v (Lander et al. 1987). The Haldane mapping

function (Haldane 1919) was used to convert recombination frequency into linkage map distance. The linkage groups were oriented and designated following Kim et al. (2005b).

In silico mapping of ESTs

Aligned sorghum genome sequence information (Sbi 1.4) for elite sorghum inbred BTx623 was made available for public use (http://www.jgi.doe.gov/sorghum) during the course of this study (Paterson et al. 2009). This aligned sorghum genome sequence information was downloaded and formatted into a database. All mapped SSR-containing EST sequences (55 markers) were searched against this database using Paracel BLAST (Altschul et al. 1990) on a Paracel High Performance Computing system. These sequences were then aligned on to the sequence-based physical map according to their hit location on each linkage group.

Results

Data mining of sorghum ESTs

In the TIGR database, about 187,282 sorghum EST sequences were available when this study was initiated. We identified 39,106 EST sequences with SSRs using the SSRIT tool. These sequences were clustered using the CAP3 program to identify 10,044 non-redundant EST sequences containing microsatellites.

A set of 1,486 non-redundant sorghum EST sequences containing microsatellites were BLAST searched against the rice genome in the GRAMENE database. From the sequence similarity search, 150 sorghum EST sequences had no hits on rice genome sequence; 180 sequences had homology with sequences on rice chromosome 1; 130 on chromosome 2, 243 on chromosome 3, 98 on chromosome 4, 115 on chromosome 5, 77 on chromosome 6, 67 on chromosome 7, 66 each on chromosomes 8 and 9, 71 on chromosome 10, 144 on chromosome 11, and 79 on chromosome 12. On each chromosome, 50 sorghum SSR-containing EST sequences with highest BLAST search score that provided coverage across the entire length of each rice chromosome were identified. Thus, 1,486 non-redundant EST sequences were BLAST searched against the rice genome to identify 600 sequences (50 per rice chromosome). Primer pairs flanking the SSR motifs were designed using Primer3. These primer pairs were designated as 'ICRISAT Sorghum EST Primers (ISEP)' (listed as electronic supplementary information S1). Among the selected 600 SSR-containing sorghum EST sequences, 470 (78%) sequences have putative annotations. The majority of these are classified as transcription factors or DNA binding proteins.

Out of the 600 sorghum EST-SSR markers developed, 63 were di-nucleotide repeats (10.5%); 425 were tri-nucleotide repeats (70.8%); 78 were tetra-nucleotide repeats (13.0%); and 34 were penta-nucleotide repeats (5.6%). The most abundant repeats among the selected EST sequences were AG and CCG. A total of 79 SSR motifs (13%) were of Class I type of SSRs (≥20 nucleotides in length) and 496 SSR motifs were Class II type of SSRs (>12 but <20 nucleotides in length), and remaining 25 SSR motifs were 10 nucleotides in length. Further details on SSR frequency and distribution in the set of EST sequences used here have been published by Jayashree et al. (2006).

Primer optimization and polymorphism assessment

PCR conditions for the 600 primer pairs were optimized using template DNA from the four inbred parental lines of the two mapping populations, and parental polymorphism was simultaneously scored. Out of 600 primer pairs, 457 (76.1%) amplified the template and 386 (84.5%) produced simple and easy to score amplification products, whereas the remaining 15.5% of template-amplifying primer pairs produced multiple fragments that were difficult to score. Most of the primer pairs that produced no amplification or gave non-specific amplification were targeting tri-nucleotide repeats. Of the 386 primer pairs that produced good amplification profiles, 133 primer pairs (34.5%) detected polymorphism between N13 and E 36-1 and 140 primer pairs detected polymorphism between BTx623 and IS 18551 (listed as electronic supplementary information S1). In both crosses, di-nucleotide repeats were found most polymorphic (30–38%), followed by tetra-nucleotide repeats in the N13 × E 36-1 cross, and penta-nucleotide repeats in the BTx623 × IS 18551 cross.

Mapping of EST-SSRs

The polymorphic markers were surveyed on 94 RILs derived from the cross N13 × E 36-1 and the segregation data were scored. A total of 55 EST-SSR markers were added (Table 1) to the existing skeleton map of this mapping population. These markers mapped to all 10 sorghum linkage groups and were well-distributed within each linkage group except in the case of SBI-06 for which the newly mapped markers clustered along the bottom of the linkage group in a region corresponding to the long arm of the sorghum chromosome. Among these 55 newly mapped EST-SSR markers, 9 mapped to SBI-09, 8 markers each mapped to SBI-02 and SBI-03, 7 mapped to SBI-01, 6 markers each mapped on SBI-04 and SBI-07, 3 markers each mapped on SBI-05, SBI-06, and SBI-10, and 2 markers mapped on SBI-08, with an average of 5.5 EST-SSR markers per linkage group. The linkage map augmented with these 55

Table 1 Details of 55 genetically mapped sorghum EST-SSR markers and their corresponding physical map locations on the aligned sorghum genome sequence

Marker	Sorghum EST sequence ID	Repeat motif	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	Approx. amplicon size in BTx623 (bp)	Alignment position on rice genome sequence	BLAST score	Linkage group	Physical map position
Xisep0101	AW679887	TG(9)	CAGATCTCCGGTTGAAGAGC	TGAGCCGAGCTCAACATACA	205	chr1: 1931190-12847810	1350	SBI-03	SBI-03
Xisep0102	AW565338	AAG(4)	CGCTGGAGTACCAGAGGAAG	AACAAAATCCGAGCCTGTTG	680	chr1: 419375-22780478	968	SBI-03	SBI-03
Xisep0107	AW744864	TGG(4)	GCCGTAACAGAGAAGGATGG	TTTCCGCTACCTCAAAAACC	205	chr1: 4946397-4946862	711	SBI-03	SBI-03
Xisep0110	**BE360507**	**CG(6)**	**GAGGGGAAGCTGGAGACC**	**TCAAGTGTACAACGCATCCAG**	**480**	**chr1: 810589-31327180**	**1821**	**SBI-09**	**SBI-03**
Xisep0138	AW923239	TA(6)	GAGATCGAGAGGCACTTTGG	CAGCGACAAGCCAATACCA	210	chr1: 40999721-41283498	1349	SBI-03	SBI-03
Xisep0202	AW679616	TGA(7)	CAACCTGTGATTGACCCATTT	AAACATGTCCAGATTCATCAAGG	195	chr2: 3957694-3958376	1235	SBI-04	SBI-04
Xisep0224	AW563693	CTG(4)	ACTGGGGTTCCTTTTCCTGT	TCCCTGATTTCCCCTCTTTT	200	chr2: 13206311-26037486	824	SBI-04	SBI-04
Xisep0228	BE356910	GAGG(3)	GACATGGCCAGCTAAGAGGA	CCATGCAGTGATCGTTGTGT	220	chr2: 7635368-30560994	746	SBI-04	SBI-04
Xisep0310	AW286133	CCAAT(4)	TGCCTTGTGCCTTGTTTATCT	GGATCGATGCCTATCTCGTC	200	chr3: 10479597-10480017	711	SBI-02	SBI-02
Xisep0314	**BE358851**	**GCC(4)**	**GTTCGACGACGACGACCT**	**GTCCTCGAACCCGACCTT**	**405**	**chr3: 1696130-28812187**	**730**	**SBI-10**	**SBI-01**
Xisep0327	AW680959	GTT(4)	CTGTTTGTGCTTGCAACTCC	TCATCGATGCAGAACTCACC	200/210	chr3: 16826707-30714737	730	SBI-01	SBI-01
Xisep0328	BE359052	AAG(4)	CATCTTCCTCCGTCAACCAT	ATCCTGCGACCCTTCTCAC	300	chr3: 20756868-33479601	1419	SBI-07	SBI-07
Xisep0422	AW285124	GCAT(3)	TGCCCGTAATTAAGCCCATA	CCCACTGCTCCAGGTAAGAA	300	chr4: 2810717-30050335	410	SBI-06	SBI-06
Xisep0439	AW745664	GAC(4)	TAGTCGAAGCAGCAGTCGTG	GTGTTCAAGTTCGACGAGCA	620	chr4: 31792724-31832792	863	SBI-07	SBI-07
Xisep0443	BE355584	GCA(7)	TCATGTACAGAGGCGACACG	AGGTCGCAACAGACACCTTC	190	chr4: 29104052-29122456	734	SBI-06	SBI-06
Xisep0449	BE357875	TCA(7)	CCGCTCATCAGTCATCACAT	ACAAAATCCATCCCACAACG	200	chr4: 365957-34743646	1808	SBI-06	SBI-06
Xisep0506	AW922528	AACG(3)	CGTGCAAGTTTGGAATTTGTC	CGGGCAGGTATAAGGTGTTG	225/430	chr5: 2076388-4402866	697	SBI-09	SBI-09
Xisep0513	AW745388	GGC(5)	GAGGGAAGAAGAAAACCCAGA	AGCCTCTCCTCCTCCTCCT	300	chr5: 803034-17097487	1053	SBI-09	SBI-09
Xisep0522	**BE361646**	**CAG(8)**	**TCATGGACCGTGTCATCG**	**GCGTACTTGCTCCACCTCTC**	**330**	**chr5: 15943072-22166226**	**1110**	**SBI-04**	**SBI-02**
Xisep0523	BE358885	TGC(4)	ACGACATGGACGACATCAGA	AACAAAAACACACGGGAAGG	220/305	chr5: 17631682-17632292	911	SBI-09	SBI-09
Xisep0550	AW746901	GA(11)	GCGGCGAGAGAGAGAGTTC	CGAGCTTGATCTTCTCGTTGA	185	chr5: 3039650-25338018	670	SBI-09	SBI-09
Xisep0607	AW923923	AGA(4)	CACGAGGATTTCACCAAACC	TGCACGTGTTCGAAATAGGA	195	chr6: 2238520-2240631	543	SBI-10	SBI-10
Xisep0608	**BE355471**	**AGA(4)**	**TTTCACCAAACCAAGCTAAGG**	**GTAGAGGCAGCCCTTCTCCT**	**205**	**chr6: 2238520-2240631**	**543**	**SBI-04**	**SBI-10**
Xisep0632	BE355933	CATG(4)	AGAGAGGAGGTCCCAAATGC	TTAAGGCCCAAACAAACTGG	195	chr6: 8908340-28128503	729	SBI-08	SBI-08
Xisep0639	BE357831	TCT(6)	TCGGACGGAGTCATCAGATA	GCCTTCGTGTCTTCTGTCCT	200	chr6: 26216943-26460093	628	SBI-10	SBI-10
Xisep0701	BE360023	TTCTT(3)	CGGTGGGAGAGACAGAGAGA	CCAATCAATACCACTCCTGTGA	200	chr7: 4061086-4061447	1228	SBI-02	SBI-02
Xisep0704	AW680310	GT(5)	CAAGTCCGTCGTGCTAGAGG	CCCTTTAATTAGCCCCAAACA	200	chr7: 5247630-5247943	908	SBI-07	SBI-07
Xisep0733	BE357808	TGTAA(5)	GTGATGCATCTCACGGACAG	GCAGCTACTGCATGCTTGTC	400	chr7: 22294109-26985575	1443	SBI-02	SBI-02
Xisep0747	BE357135	TCC(5)	AGGCAGCCTGCTTATCACAA	ACAAGCTCAGGTGGGTGGT	210	chr7: 372151-26726597	897	SBI-02	SBI-02
Xisep0809	AW747322	TATG(4)	GGAAACTCTTGTGGGTTGGA	TTGACCTCTCTACAAATGATCCAC	200	chr8: 1022442-1022570	221	SBI-08	SBI-08
Xisep0824	AW564458	CCG(4)	TCCTGAAAGAAACGCACACA	GAGGAGGGTGTGGAGGTGTA	200	chr8: 13585814-13586034	428	SBI-03	SBI-03
Xisep0829	BE125697	AG(6)	CGCTGCCAAAATCTAAGCTC	CACGGTGGTCACATCAGAAG	205/243	chr8: 27738244-27738554	1225	SBI-07	SBI-07
Xisep0831	BE360943	AAAAG(3)	TCCATGACCTTGAGGAGGAG	TTGAAGCAGGACAACACACC	200	chr8: 24781619-24781723	943	SBI-07	SBI-07
Xisep0839	BE357549	ATTAC(4)	TACGCATAGCGCCTTTCAAT	ATTTCATTATGCCGGTCTCG	200	chr8: 27711931-27712047	378	SBI-01	SBI-01
Xisep0841	BE358373	GCA(10)	TAGGAATGACGACACCACCA	CAAAGGCAAGGGTTTTGCTA	210	chr8: 24665284-24665519	299	SBI-02	SBI-02

Table 1 continued

Marker	Sorghum EST sequence ID	Repeat motif	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	Approx. amplicon size in BTx623 (bp)	Alignment position on rice genome sequence	BLAST score	Linkage group	Physical map position
Xisep0843	BE357146	AGCTG(3)	CCCAAACATTCCCACGTAAC	GTGAACAGAGGAGGCAGAGG	200	chr8: 2991852-24982703	993	SBI-03	SBI-03
Xisep0844	BE356742	CGT(4)	GTGTTCAAGTTCGACGAGCA	TAGTCGAAGCAGCAGTCGTG	600	chr8: 1641413-25156453	941	SBI-07	SBI-07
Xisep0938	AW747241	TGGGT(6)	TGCTGTTCTTGAACGTGTTTG	TTTTGCACAAAGTTGCGTGT	225	chr9: 17210486-17210944	779	SBI-02	SBI-02
Xisep0948	BE358472	TA(5)	AGGCCGAATCACAATAATGG	AGTGCATGAACAGGGCATC	205	chr9: 20568536-20568726	297	SBI-04	SBI-04
Xisep0949	AW564537	GCA(5)	CAGTGCCAATAAGCTCGTCTC	CATCGATCTCTGCTTCTGCTT	100	chr9: 19365772-19365933	271	SBI-01	SBI-01
Xisep1008	**AW285562**	**CAG(7)**	**GATGCGCAAGCAGAACAAG**	**CAGCAATGGAAATAGCTCAGG**	**340**	**chr10: 2491921-13066504**	**632**	**SBI-09**	**SBI-01**
Xisep1013	BE360610	CGG(7)	CGGTTACGGCGGATTATTAC	ATGGTGGCGATGCAGACTA	220	chr10: 4414394-16362660	402	SBI-02	SBI-02
Xisep1014	AW283055	GT(5)	ACCGCCGACGTCATAGTAAG	GGCAGTAACATAGCATCCATCA	240	chr10: 13864191-13864421	391	SBI-09	SBI-09
Xisep1025	**AW672031**	**GCG(5)**	**ACCTTCTCGTCCTCGTCCTC**	**AGAACATGACCGGATCGAAG**	**490**	**chr10: 8526382-21813965**	**931**	**SBI-02**	**SBI-01**
Xisep1028	AW747772	GCA(4)	CAGCGACCATGAGGATGAC	TGGCATGCATCAAACAAGAT	200	chr10: 11501109-20941969	638	SBI-01	SBI-01
Xisep1031	BE355884	CGG(6)	TGCTCCTGCCTCGTTCTC	TAGTCCTCGGTCGACTCCAT	690	chr10: 5554237-21765954	326	SBI-03	SBI-03
Xisep1032	BE356267	TTCAG(3)	GCAAGCTCTACGGGATCTTC	GCAGCTGGAAAATAATCGAAA	200	chr10: 14058773-21609683	1532	SBI-01	SBI-01
Xisep1039	AW287556	CCTG(5)	GTGGATTCAAATCCGCTGAC	GGCAATTTGGCAAGCAAT	200	chr10: 13165407-17730031	492	SBI-01	SBI-01
Xisep1047	AW285966	TGCGT(4)	TCTTTGCCCTCCCGCTAT	TCCATGAAATGAGCAAACGA	160	chr10: 18901923-18902174	235	SBI-03	SBI-03
Xisep1107	BE125211	GCA(6)	GGATAATCTGCAGGCGACTT	CCATCTGCTGCTCTGACTTG	210	chr11: 5341585-5362168	830	SBI-05	SBI-05
Xisep1111	AW286876	CGAG(3)	CCTCCTCTCCTCATCCCTCT	ATGGGTAGCGGGTTTCTTG	220	chr11: 1392730-1393672	711	SBI-05	SBI-05
Xisep1128	**AW565976**	**AT(6)**	**GGCGGGAAAAAGTTCCTTTA**	**CGCACACCCATTTCAATTC**	**220**	**chr11: 14249874-23849881**	**793**	**SBI-09**	**SBI-05**
Xisep1133	AW564774	CCA(5)	CGATGCAGCTCCAACTCATA	GTGTATGTCGCCGAAGTGG	220	chr11: 2943934-21608323	350	SBI-05	SBI-05
Xisep1213	BE358492	GAA(5)	AGGTCAGCGTCTTGCAATCT	ACAAATTGAAAGGGCGAGAG	202	chr12: 1658936-21046139	315	SBI-01	SBI-01
Xisep1241	BE360849	AGCTG(5)	GAGGGCGAGACAGAGGAGAT	CTACCTTTGAGCCCACCGTA	190	chr12: 24812252-24812523	288	SBI-09	SBI-09

Markers conflicting in physical and linkage map positions are highlighted in bold

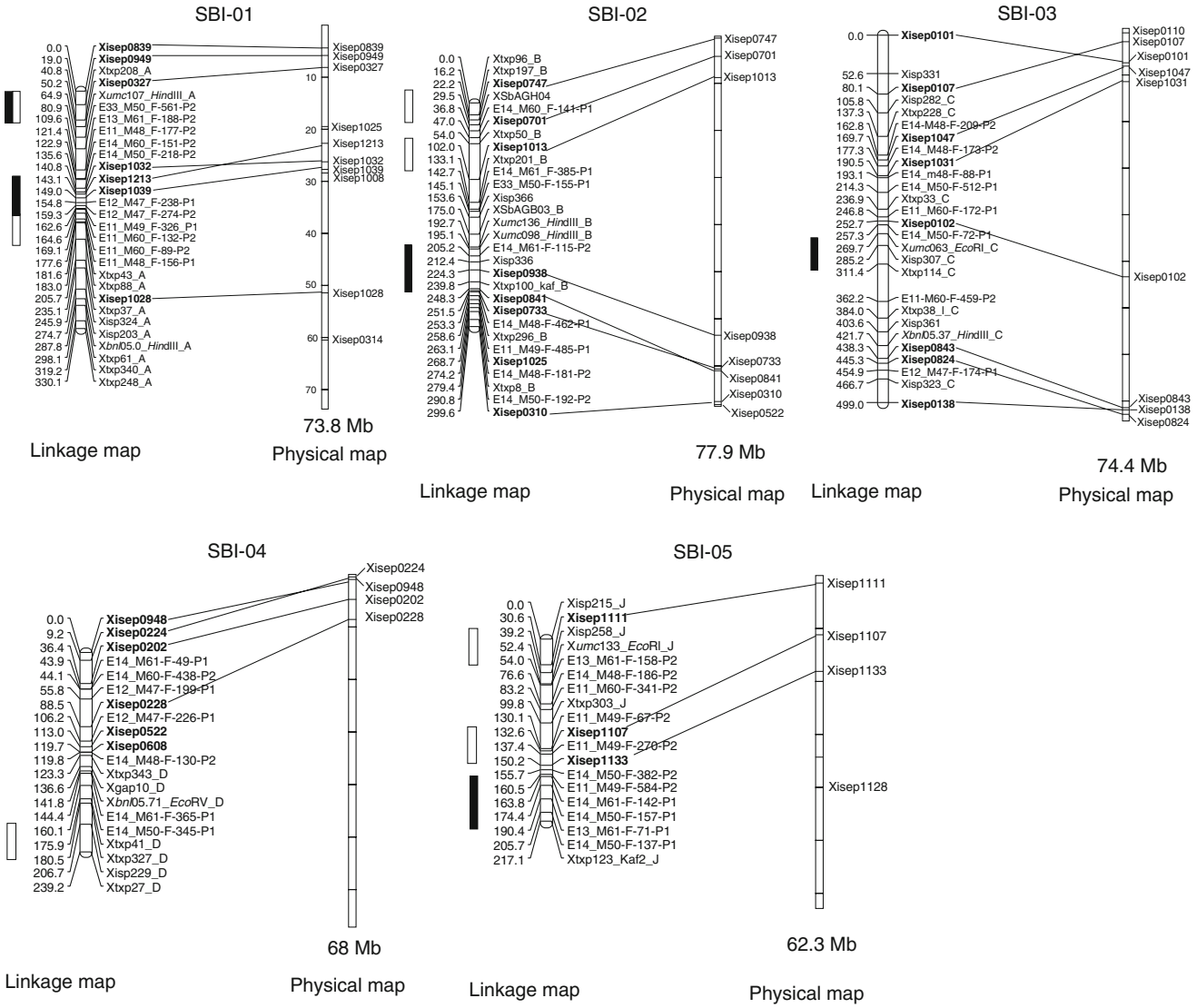

Fig. 1 Distribution of EST-SSR markers across sorghum linkage groups. Linkage group nomenclature follows Kim et al. (2005b). Newly added markers are highlighted in bold font. *Striga* resistance QTL and stay-green QTL

newly mapped loci spanned 2,838 cM (Fig. 1). Markers such as *Xisep0841*, *Xisep0733*, and *Xisep0938* on SBI-02, and *Xisep0138*, *Xisep0824*, and *Xisep0843* on SBI-03, mapped to the vicinity of previously reported stay-green QTLs (Haussmann et al. 2002b) whereas *Xisep0747*, *Xisep0701*, and *Xisep1013* on SBI-02 mapped in the vicinity of previously reported *Striga* resistance QTLs (Haussmann et al. 2004).

In silico mapping of SSR containing ESTs

All genetically mapped SSR-containing ESTs were then mapped physically by BLAST search against the aligned sorghum genome sequence. These markers were then aligned on to the sorghum physical map (Fig. 1). Physical map positions were essentially as expected from the genetic linkage analysis of these markers, except in case of *Xisep0110*, *Xisep0314*, *Xisep0522*, *Xisep0608*, *Xisep1008*, *Xisep1025*, and *Xisep1128*. Physical and linkage map positions for all 55 mapped markers are listed in Table 1.

Discussion

The nature and frequency of SSRs in sorghum EST collection has been comprehensively discussed in Jayashree et al. (2006). Generally, EST-derived SSR markers are found to be less polymorphic than genomic SSRs. In this study, only 28% of the EST-SSR markers developed were polymorphic among the four sorghum genotypes initially screened. A similar percentage of polymorphism (25%) was reported for EST-SSRs in durum wheat (Eujayl et al. 2002), whereas

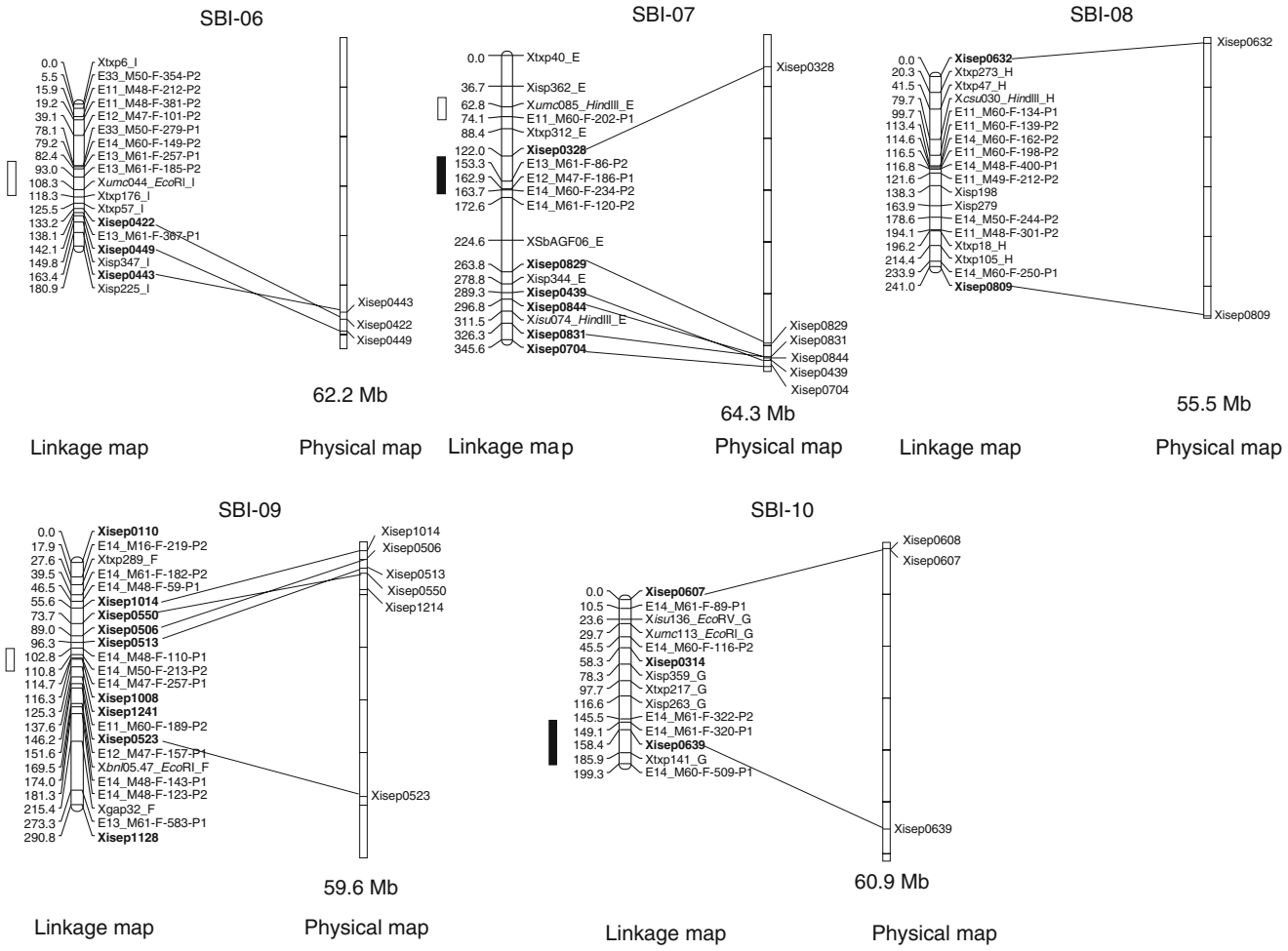

Fig. 1 continued

Thiel et al. (2003) reported 8–54% polymorphism on three different mapping population parental line pairs in barley. Di-nucleotide repeats were more polymorphic (listed as electronic supplementary information S1) than other repeat classes, as reported by Thiel et al. (2003) in barley. In our study, a total of 169 EST-SSR markers (28%) detected polymorphism between parental lines of two mapping populations. Among these, 104 EST-SSR markers were polymorphic in both populations whereas 36 and 29 detected polymorphism specific to the shoot fly and the *Striga* resistance mapping populations, respectively.

Sequence-based alignment studies accelerate the filling of gaps on linkage maps of related species and can be effectively used for comparative genome mapping across species (Klein et al. 2003). When searching the homology of 1,486 non-redundant sorghum SSR-containing EST sequences against the rice genome, the highest alignment scores were obtained for those aligning to the long arms of rice linkage groups (listed as electronic supplementary information S1). Rice chromosome 3 had the highest number of hits followed by chromosomes 1 and 11. In the sequence similarity search, 150 of these sorghum EST sequences did not have any hit on the rice genome. Klein et al. (2003) also found that 10% of the sorghum ESTs studied have no homologs in the rice genome. Interestingly, this agrees well with the observation that 7% of the predicted genes from the aligned 8× shotgun sequence of sorghum inbred BTx623 have no apparent homologs in *Arabidposis*, rice or poplar (Paterson et al. 2009).

After incorporating the new markers, the total map distance for the 94-entry (N13 × E 36-1)-based RIL population was 2,838 cM, which is similar to that reported by Tao et al. (2000) using 152 RILs and 306 markers. However, the present map is longer than sorghum linkage maps previously reported by several authors (Haussmann et al. 2002a; Pereira et al. 1994; Chittenden et al. 1994). The variation in map length of different mapping populations is generally attributed to differences in population size, the type of population, and the genetic distance between the parents, which strongly affect observed recombination rates, and thus, genetic distances (Menz et al. 2002). A comparison of the present map with that of Haussmann et al. (2002a), which

was based on a larger sample of progenies from the same mapping population, revealed that 94 out of 113 (83%) common markers had the same linear order in both maps. Some possible rearrangements were detected, which mostly involved alternative orders for closely linked loci. Such rearrangements are likely due to the use of a sub-sample of 94 lines from this mapping population in the current study. The only major change observed in the order of markers was in the upper part of linkage group SBI-04, inverting the order of two markers separated by 9 cM. These EST-SSR markers, however, filled gaps or marker-rare regions in the existing linkage map of sorghum targeted based on rice–sorghum sequence similarity.

The newly developed EST-SSR markers have shown practical significance by virtue of their positions. For example, *Xisep0138*, *Xisep0843*, and *Xisep0824*, which were mapped to SBI-03, are located in the vicinity of a stay-green QTL reported by Haussmann et al. (2002b) (Fig. 1). Three of the markers mapped to SBI-05 (*Xisep1111*, *Xisep1107*, and *Xisep1133*) are located in close proximity to *Striga* resistance QTLs (Haussmann et al. 2004). These markers are currently being used by ICRISAT and our national program partners in the introgression of stay-green and *Striga* resistance QTLs into a wide range of farmer-preferred sorghum genotypes through marker-assisted backcrossing. Two markers [*Xisep0107* (SBI-03) and *Xisep0310* (SBI-02)] were selected based on their linkage map positions and used by the Generation Challenge Programme (GCP) in large-scale SSR-based diversity analysis of over 3,000 sorghum accessions in a global composite germplasm collection. Across this very broad range of wild and cultivated sorghums, a total of nine alleles were noted for *Xisep0310* and six alleles for *Xisep0107* (data not shown). Such small allele numbers with EST-SSRs in such a large and diverse set of germplasm, as compared to genomic SSRs agrees with expectations that EST-SSRs are less polymorphic than genomic SSRs, as the former are derived from conserved and expressed regions of the genome. Although this makes EST-SSRs less useful than genomic SSRs for fingerprinting purposes, this lower level of variability is helpful in diversity analysis, in which relationships between accessions are much more difficult to assess when there are large numbers of rare alleles. Hence, these EST-SSRs have potential for use in assessing functional diversity among different genotypes, as well as use as flanking markers for foreground selection in marker-assisted breeding programs.

In silico mapping of the ESTs (from which these new SSRs were developed) on the aligned sorghum genome sequence (Paterson et al. 2009) gave very similar positions as the conventional linkage analysis, both in terms of chromosome arm location and order. The order of these EST-SSRs from the *in silico* mapping agrees with the linkage map for all ten sorghum linkage groups provided that the latter are oriented with the short chromosome arm at the top, as per the suggestion of Kim et al. (2005a). Only seven (12.7%) out of 55 mapped EST-SSR markers added to the N13 × E 36-1-based skeleton map were mapped *in silico* to sorghum chromosomes other than those expected. Some of these differences are likely because the ESTs have been drawn of multi-gene families with members distributed on more than one chromosome as a result of ancestral genome duplication events or transposition. For example, *Xisep1025* mapped *in silico* to SBI-01 where as it is mapped to SBI-02 based on linkage analysis. However, the sorghum EST sequence (which had a best hit on rice chromosome 1) from which the primer pair for *Xisep1025* was derived also has a low e-value hit on SBI-02. This may be due to presence of duplicate loci on SBI-01 and SBI-02. Recently, 5,012 genomic SSRs were mapped *in silico* on to the aligned sorghum genome sequence assembly (Yonemaru et al. 2009).

The sorghum genome sequencing project has revealed that predicted gene density in sorghum is much higher at the ends of each chromosome and that heterochromatic regions near the centromere are essentially devoid of predicted genes (Paterson et al. 2009). *In silico* positions for all mapped EST-SSRs are located at the ends of sorghum chromosome arms, away from the centromeric regions, in agreement with the predicted genomic distribution of sorghum genes. The *in silico*-predicted physical map positions of candidate markers (including EST-SSRs that have not yet been mapped) can help to choose markers to fill gaps and better saturate linkage maps. Thus the availability of genome sequences and other marker information can 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.

On-going and future sequencing projects will contribute many more genomic and genic sequences to publicly available databases, facilitating development of different types of markers at lower cost. As economically important grasses have more conserved regions across taxa than found in other crop lineages, EST-SSRs should become an extremely powerful tool for better understanding of relationships between the grass species and for genetic mapping. Our initial attempt to develop EST-SSRs based on sorghum–rice synteny was successful in filling important gaps in the existing sorghum linkage map. We propose to investigate further the large number of SSR-containing sorghum EST sequences to develop a dense linkage map based on these functional markers. The newly developed markers in hand, which are tightly linked to genomic regions controlling *Striga* resistance and the stay-green component of terminal drought tolerance, can be used in functional

analysis of these traits. These markers are being used by ICRISAT, along with the newly reported CISP markers (Feltus et al. 2006b), for mapping of sorghum resistance to shoot fly, stem borers, and grain mold. Through this study, we have demonstrated the value of a comparative sequence similarity approach for targeted development of PCR-compatible molecular markers for practical applications in crop genetics and breeding. It has previously been reported that more than half (57%) of EST-SSR markers developed from sorghum ESTs show cross-species transferability (Wang et al. 2005), hence these markers can also be successfully used in other related grass species for which insufficient numbers of PCR-compatible markers are available. Primer pairs for non-redundant sorghum EST-SSRs are available from our database (http://intranet.icrisat.org/gt1/ssr/ssrdatabase.html).

Acknowledgments The research fellowship provided to PR by the Council of Scientific and Industrial Research (CSIR), New Delhi, India is greatly acknowledged. KB acknowledges the financial support provided by the Ethiopian Government and ICRISAT. Work reported here was partially funded by the CGIAR Generation Challenge Program (GCP).

References

Agrama HA, Widle GE, Reese JC, Campbell LR, Tuinstra MR (2002) Genetic mapping of QTLs associated with green bug resistance and tolerance in *Sorghum bicolor*. Theor Appl Genet 104:1373–1378

Ahn S, Anderson JA, Sorrells ME, Tanksley SD (1993) Homoeologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 241:483–490

Altschul SF, Gish W, Miller W, Myers WE, Ipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Antonopoulou G, Gavala HN, Skiadas IV, Angelopoulos K, Lyberatos G (2008) Biofuels generation from sweet sorghum: fermentative hydrogen production and anaerobic digestion of the remaining biomass. Bioresource Technol 99:110–119

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

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

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

Bowers JE, Abbey C, Anderson S, Chang C, Draye X et al (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

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT et al (2005) Comparative physical mapping links conservation of micro-synteny to chromosome structure and recombination in grasses. Proc Nat Acad Sci (USA) 102:13206–13211

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich 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

Chen M, Presting G, Barbazuk WB et al (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

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

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

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35:3–15

Draye X, Lin Y-R, Qian X-Y, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren SX, Wing RA, Paterson AH (2001) Towards integration of comparative genetic, physical, diversity, and cyto-molecular maps for grasses and grains, using the sorghum genome as a foundation. Plant Physiol 125:1325–1341

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

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Feltus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham S, Klein PE, Brown PJ, Paterson AH (2006a) Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. Theor Appl Genet 112:1295–1305

Feltus FA, Singh HP, Lohithaswa HC, Schilze SR, Silva TD, Paterson AH (2006b) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. Plant Physiol 140:1183–1191

Folkertsma RT, Sajjanar GM, Reddy BVS, Sharma HC, Hash CT (2003) Genetic mapping of QTL associated with sorghum shootfly (*Atherigona soccata*) resistance in sorghum (*Sorghum bicolor*). In: XI International Plant and Animal Genome conference, San Diego, CA, USA, W65 (http://www.intl-pag.org/11/abstracts/P5d_P462_XI.html)

Goff SA, Ricke D, Lan TH, Presting G, Wang R et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Klein R, Mullet J (2007) Sorhum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J Exp Bot 58:327–338

Hart GE, Schertz KF, Peng Y, Syed NY (2002) Genetic mapping of *Sorghum bicolor* (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet 103:1232–1242

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 stay-green in two sorghum recombinant inbred populations. Theor Appl Genet 106:133–142

Haussmann BIG, Hess DE, Omanya GO, Folkertsma RT, Reddy BVS, Kayentao M, Welz HG, Geiger HH (2004) Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in

two recombinant inbred populations of sorghum. Theor Appl Genet 109:1005–1016
Hicks C, Tuinstra MR, Pedersen JF, Dowell FE, Kofoid KD (2002) Genetic analysis of feed quality and seed weight of sorghum inbred lines and hybrids using analytical methods and NIRS. Euphytica 127:31–40
Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877
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
International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800
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/
Katsar C, Paterson AH, Teetes GL, Peterson GC (2002) Molecular analysis of sorghum resistance to the greenbug (Homoptera: Aphididae). J Econ Ent 95:448–457
Kebede H, Subudhi PK, Rosenow DT, Nguyen HT (2001) Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). Theor Appl Genet 103:266–276
Kim J-S, Islam-Faridi MN, Klein PE, Stelly DM, Price HJ, Klein RR, Mullet JE (2005a) Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. Genetics 171:1963–1976
Kim J-S, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005b) Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169:1169–1173
Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morshige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res 10:789–807
Klein RR, Rodriguez-Herrera R, Schlueter JA, Klein PE, Yu ZH, Rooney WL (2001) Identification of genomic regions that affect grain-mould incidence and other traits of agronomic importance in sorghum. Theor Appl Genet 102:307–319
Klein PE, Klein RR, Vrebalov J, Mullet JE (2003) Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. Plant J 34:605–621
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
Kresovich S, Barbazuk B, Bedell JA, Borrell A, Buell AR et al (2005) Towards sequencing the sorghum genome. A U.S. National Science foundation-sponsored workshop report. Plant Physiol 138:1898–1902
Lander E, Green P, Abrahamson J, Barlow A, Daley M, Lincoln S, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181
Lijavetsky D, Martinez MC, Carrari F, Hopp HE (2000) QTL analysis and mapping of pre-harvest sprouting resistance in *Sorghum*. Euphytica 112:125–135
Lin YR, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae in reference to an interspecific sorghum population. Genetics 141:391–411
Mace ES, Buhariwalla HK, Crouch JH (2003) A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Rep 21:459a–459h
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 Genomics 9:26
Magalhaes JV, Garvin DF, Wang Y, Sorrells ME, Klein PE, Schaffert RE, Li L, Kochian LV (2004) Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae. Genetics 167:1905–1914
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 Bio 48:483–499
Nagaraj N, Reese JC, Tuinstra MR, Klein PE, SanMiguel P, Chen C-P, Li J, Devos KM, Schertz K, Dunkle L, Bennetzen JL (2005) Molecular mapping of sorghum genes expressing tolerance to damage by greenbug (Homoptera: Aphididae). J Econ Entomol 98:595–602
Nagy ER, Lee TC, Ramakrishna W, Xu Z, Klein PE, SanMiguel P, Cheng CP, Li J, Devos KM, Schertz K, Dunkle L, Bennetzen JL (2007) Fine mapping of the *Pc* locus of *Sorghum bicolor*, a gene controlling the reaction to a fungal pathogen and its host-selective toxin. Theor Appl Genet 114:961–970
Oh B-J, Frederiksen RA, Magill CW (1996) Identification of RFLP markers linked to a gene for downy mildew resistance (*Sdm*) in sorghum. Can J Bot 74:315–317
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, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556
Peng Y, Schertz KF, Cartinhour S, Hart GE (1999) Comparative genome mapping of *Sorghum bicolor* (L.) Moench using an RFLP map constructed in a population of recombinant inbred lines. Plant Breed 118:225–235
Pereira MG, Lee M (1995) Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet 90:380–388
Pereira MG, Lee M, Bramel-Cox P, Woodman W, Doebley J, Whitkus R (1994) Construction of an RFLP map in sorghum and comparative mapping in maize. Genome 37:236–243
Ragab RA, Dronavalli S, Maroof MAS, Yu YGL (1994) Construction of a sorghum RFLP linkage map using sorghum and maize DNA probes. Genome 37:590–594
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
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
Subudhi PK, Nguyen HT (2000) Linkage group alignment of sorghum RFLP maps using a RIL mapping population. Genome 43:240–249
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
Tao YZ, Jordan DR, Henzell RG, McIntyre CL (1998a) Construction of a genetic map in sorghum RIL population using probes from different sources and its comparison with other sorghum maps. Aust J Agric Res 49:729–736
Tao YZ, Jordan DR, Henzell RG, McIntyre CL (1998b) Identification of genomic regions for rust resistance in sorghum. Euphytica 103:287–292

Tao YZ, Henzell RG, Jordan DR, Butler DG, Kelly AM, McIntyre CL (2000) Identification of genomic regions associated with stay-green in sorghum by testing RILs in multiple environments. Theor Appl Genet 100:1225–1232

Tao YZ, Hardy A, Drenth J, Henzell RG, Franzmann BA, Jordan DR, Butler DG, McIntyre CL (2003) Identification of two different mechanism s for sorghum midge resistance trough QTL mapping. Theor Appl Genet 107:116–122

Taramino G, Tarchini R, Ferrario S, Lee M, Pé ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95:66–72

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *Adh*1-*Adh*2 region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381–391

Tegelstrom H (1992) Detection of mitochondrial DNA fragments. In: Hoelzel AR (ed) Molecular genetic analysis of populations: a practical approach. IRL Press, Oxford, pp 89–114

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1996) Identification of quantitative trait loci associated with pre-flowering drought tolerance in sorghum. Crop Sci 36:1337–1344

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1997) Genetic analysis of post-flowering drought tolerance and components of grain development in *Sorghum bicolor* (L.) Moench. Mol Breed 3:439–448

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotech 23:48–55

Ventelon M, Due M, Garsmeur O, Doligex A, Ghesquiere A, Lorieux M, Rami JF, Glaszmann JC, Grivet L (2001) A direct comparison between the genetic maps of sorghum and rice. Theor Appl Genet 102:379–386

Wang ML, Barkley NA, Yu JK, Dean RE, Newman ML, Sorrells ME, Pederson GA (2005) Transfer of simple sequence repeat (SSR) markers from major cereal crops to minor grass species for germplasm characterization and evaluation. Plant Genet Resour 3:45–57

Wu YQ, Huang Y, Porter DR, Tauer CG, Holloway L (2007) Identification of a major quantitative trait locus conditioning resistance to greenbug biotype E in sorghum PI 550610 using simple sequence repeat markers. J Econ Entomol 100:1672–1768

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 stay-green 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 1–7. doi:10.1093/dnares/dsp005

Yu J, Hu S, Wang J, Wong GK, Li S et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92