

SHORT COMMUNICATION

Development of Genome-wide Simple Sequence Repeat Markers Using Whole-genome Shotgun Sequences of Sorghum (*Sorghum bicolor* (L.) Moench)

JUN-ICHI Yonemaru^{1,*}, TSUYU Ando², TATSUMI Mizubayashi², SHIGEMITSU Kasuga³, TAKASHI Matsumoto¹, and MASAHIRO Yano¹

National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan¹; Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki 305-0854, Japan² and Education and Research Center of Alpine Field Science, Faculty of Agriculture, Shinshu University, Minamiminowa, Nagano 399-4598, Japan³

(Received 22 December 2008; accepted 13 March 2009; published online 10 April 2009)

Abstract

Simple sequence repeat (SSR) markers with a high degree of polymorphism contribute to the molecular dissection of agriculturally important traits in sorghum (*Sorghum bicolor* (L.) Moench). We designed 5599 non-redundant SSR markers, including regions flanking the SSRs, in whole-genome shotgun sequences of sorghum line ATx623. (AT/TA)_n repeats constituted 26.1% of all SSRs, followed by (AG/TC)_n at 20.5%, (AC/TG)_n at 13.7% and (CG/GC)_n at 11.8%. The chromosomal locations of 5012 SSR markers were determined by comparing the locations identified by means of electronic PCR with the predicted positions of 34 008 gene loci. Most SSR markers had a similar distribution to the gene loci. Among 970 markers validated by fragment analysis, 67.8% (658 of 970) markers successfully provided PCR amplification in sorghum line BTx623, with a mean polymorphism rate of 45.1% (297 of 658) for all SSR loci in combinations of 11 sorghum lines and one sudangrass (*Sorghum sudanense* (Piper) Stapf) line. The product of 5012 and 0.678 suggests that ~3400 SSR markers could be used to detect SSR polymorphisms and that more than 1500 (45.1% of 3400) markers could reveal SSR polymorphisms in combinations of *Sorghum* lines.

Key words: sorghum (*Sorghum bicolor* (L.) Moench); simple sequence repeat (SSR); fragment analysis; genome-wide

Sorghum (*Sorghum bicolor* (L.) Moench) is the world's fifth most important cereal crop, after wheat, rice, maize and barley¹ and was grown on 43 million hectares in 2004 (<http://faostat.fao.org/>). There is a great demand for sorghum for human food, livestock feed and green manure production. As in other crop species, it is necessary to improve

traits with economic value, including yield potential, sugar content and disease resistance, to enhance sorghum's potential as a crop. To dissect the morphological and physiological traits of sorghum at a genetic level, many types of molecular markers have been developed, including restriction-fragment-length polymorphisms (RFLPs),^{2–4} amplified-fragment-length polymorphisms⁵ and simple sequence repeats (SSRs).^{6–11} SSR markers are mostly codominant, are readily amplified by polymerase chain reaction (PCR) and are effective at detecting genotype variation caused by a high degree of polymorphism.

Edited by Satoshi Tabata

* To whom correspondence should be addressed. Tel./Fax. +81 29-838-7135. E-mail: yonemaru@affrc.go.jp

© The Author 2009. Kazusa DNA Research Institute.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

However, only a few hundred sorghum SSR markers are publicly available for use in fine mapping and map-based cloning of genes of interest.

Among the five major cereal crops, the genome of sorghum is the second smallest (750 Mb) after that of rice (440 Mb) and is between one-third and one-quarter the size of the maize genome (2500 Mb).¹² Sorghum is more closely related to maize and sugarcane than to rice, and it shared a common ancestor as recently as 18–25 million years ago with maize¹³ and 10 million years ago with sugarcane.¹⁴ Owing to its small genome size and evolutionary divergence, sorghum is a potentially important model plant among the panicoid grasses. In sorghum, whole-genome shotgun (WGS) sequencing by means of methylation filtration has tagged 96% of the sorghum genes,¹⁵ and recently the whole genome was sequenced and annotated.¹⁶ These sequence data were produced by the U.S. Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) in collaboration with the research community. The most recent versions of the Sbi1 and Sbi1_4 annotations use pseudo-chromosome alignment (JGI Phytozome *Sorghum bicolor* website: <http://www.phytozome.net/sorghum>). The availability of sorghum genomic sequences now provides an opportunity to extend the bank of SSR markers. The sorghum genome sequence project identified ~71 000 SSRs in the genome;¹⁶ however, no information on the SSR motifs and primers is available.

We have developed a new set of SSR markers to facilitate the genetic and molecular dissection of sorghum genes that encode traits with economic value, including quantitative traits. This was achieved by using the following procedures: (i) design of

primer pairs in regions flanking the SSR motifs from WGS sequences; (ii) mapping of the genomic positions of designed SSR primer pairs by means of electronic PCR (ePCR)¹⁷ and comparison of the positions of the new SSR markers with gene loci predicted using the Sbi1_4 annotations and (iii) experimental validation of representative SSR markers by means of fragment analysis among 11 sorghum lines and one sudangrass line.

The screening of 570 794 WGS sequences (accession numbers CL147592–CL197752, CC058553–CC059980, BZ329127–BZ342789, BZ342901–BZ352342, BZ365856–BZ368372, BZ369686–BZ370012, BZ421595–BZ424357, BZ625682–BZ629992, BZ779555–BZ781928, CW020594–CW502582 and CW512190–CW514008) from the ATx623 line of sorghum for di-, tri-, tetra- and pentanucleotide repeats provided 11 684 sequences suitable for the design of SSR markers using the SSRIT software.¹⁸ To minimize duplication, the 11 684 sequences were screened for redundancy by comparisons among the primer sequences and by means of BLASTN analysis against genomic sequences. Of the 570 794 sequences, 0.98% (5599) contained unique SSR markers that were non-redundant and non-overlapping in sorghum chromosomes. These sequences are referred to as SSR markers and are provided in Supplementary Table S1. Primer sets for SSR markers were automatically designed by using the Perl script *srchssr2.pl*¹⁹ to control the Primer3 core program.²⁰ All parameters of *srchssr2.pl* were set at the default values.

An analysis of the association between SSR motifs and the rate of polymorphism is important for the development of effective SSR markers. Table 1

Table 1. Distributions of the 10 most common SSR motifs in the data set of 5599 newly developed SSR markers (which are described in Supplementary Table S1)

Motif	Chromosome											All	Frequency	Mean number of repeats
	1	2	3	4	5	6	7	8	9	10	Not determined ^a			
AT/TA	848	609	644	499	411	426	378	364	351	368	1071	5969	26.1	6.74
AG/TC	672	573	585	490	257	372	332	251	344	371	436	4683	20.5	6.78
AC/TG	445	367	412	324	186	252	186	177	220	213	362	3144	13.7	4.74
CG/GC	425	295	336	355	109	226	158	109	226	217	249	2705	11.8	3.31
CT/GA	190	145	151	117	73	93	88	80	93	104	145	1279	5.6	3.27
CA/GT	160	115	142	112	41	86	69	72	65	83	125	1070	4.7	3.12
CGC/GCG	131	96	100	102	51	48	52	72	92	84	58	886	3.9	4.23
CGT/GCA	134	95	110	100	44	58	61	62	56	88	57	865	3.8	4.49
CGA/GCT	103	88	97	97	41	51	73	57	73	82	51	813	3.6	4.42
GGA/CCT	121	71	108	99	34	51	55	52	75	67	40	773	3.4	4.69
Sum	3229	2454	2685	2295	1247	1663	1452	1296	1595	1677	2594	22 187	96.9	

^aPosition of SSR marker could not be determined by means of ePCR program.

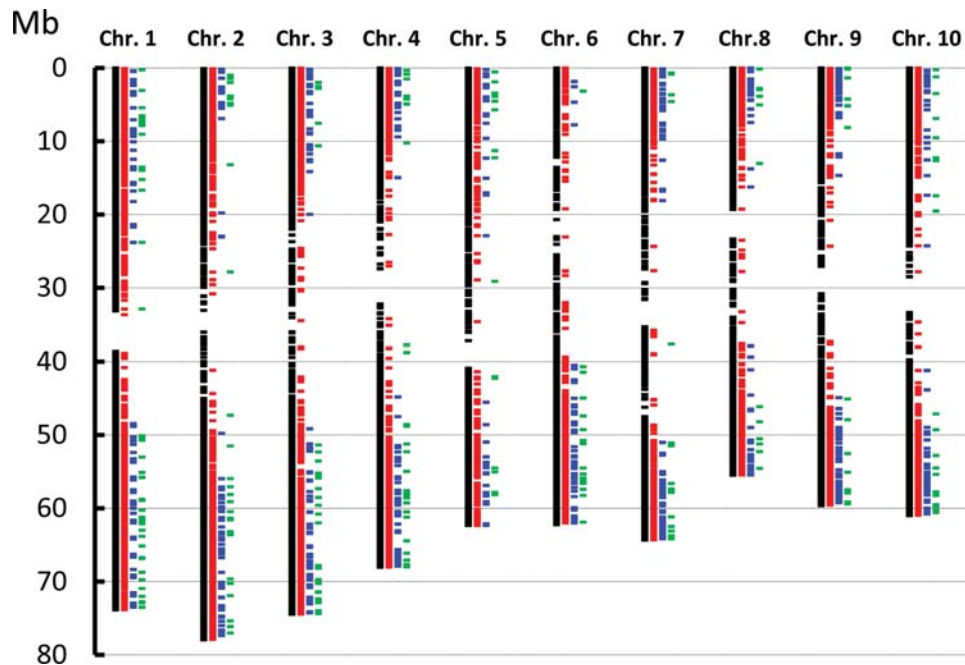


Figure 1. Distribution of SSR markers within the sorghum chromosomes. Bars at each chromosome represent the positions of the 34 008 gene loci,¹⁶ SSR markers developed in this study, polymorphic SSR markers without an AT/TA motif detected by means of fragment analysis and published SSR markers^{6–9,11} from left to right. Positional information for loci was extracted from within the Sbi1_4 annotation database, and all sequence data sets for the chromosomes were used only for SSR mapping by means of ePCR.

summarizes the relative frequency of the 10 most common SSR motifs among the set of 5599 SSR markers. The most common dinucleotide motifs were AT/TA (26.1%), AG/TC (20.5%), AC/TG (13.7%) and CG/GC (11.8%). McCouch et al.²¹ reported that AT-rich motifs in rice have a larger number of repeats and longer repeat tracks than other dinucleotide motifs and were associated with high rates of polymorphism. In our study, the most repetitive SSR motif with a perfect repeat was (TAT)₆₅ in the AT-rich SSR marker SB4593. Moreover, the mean repeat number for the (AT/TA)_n motif with three or more repeats was 6.74, which is approximately twice the mean repeat number of 3.31 calculated for the (CG/GC)_n motif. We therefore concluded that sorghum SSR markers with an AT-rich motif would show higher polymorphism than SSR markers with other dinucleotide motifs.

WGS sequence analysis by means of methylation filtration¹⁵ is likely to lead to the design of SSR markers from hypomethylated or low-copy regions of the sorghum genome. Bedell et al.¹⁵ revealed a 200% increase in the number of SSRs in WGS sequences identified by means of methylation filtration compared with the number identified without filtration and an increase in the overall proportion of GC-rich trinucleotide sequences. Similarly, we found that the most frequent trinucleotide motif was GC-rich (CGC/GCG)_n and that the rate of appearance of this motif was 3.9%. GC-rich and

trinucleotide motifs have also been reported as highly frequent motifs among the documented SSRs found in expressed sequence tags of wheat,²² barley,²³ sugarcane,²⁴ perennial ryegrass²⁵ and maize.²⁶ GC-rich and trinucleotide-based microsatellites are most likely to be derived from the coding region of the genome.

Map positions for 89.5% (5012 of 5599) of the markers were provided by means of ePCR. Failures to obtain map positions for some 10.5% of the markers may have resulted from the design of the primers on the basis of the WGS sequences of the male-sterile line ATx623, which would not have been mapped when the Sbi1 data derived from the male sterility-maintenance line BTx623 were used for ePCR analysis. To identify target genes and analyze quantitative trait loci in detail, it is necessary to prioritize the development of markers in gene-rich regions. The distribution of our SSR markers appears to be preferentially located at positions with 34 008 gene loci (Fig. 1), and it can be attributed to WGS sequences covering 96% of the sorghum genes over 65% of the genome length.¹⁵ The low number of designed SSR markers near the centromeres of the chromosomes is consistent with the low annotation rate for genes at these locations (Fig. 1). It is possible that these loci correspond to methylated or multi-copied genes.

The ePCR mapping of 317 SSR markers established in previous studies^{6–9,11} identified the chromosomal

locations of 260 of the markers, of which 21 had 50 bp or more of their sequences in common (Supplementary Table S2). This result shows that 239 of the previous SSR markers have different positions within the sorghum genome. A comparison of the distribution of the 5012 markers identified here with the distribution of the 239 non-redundant markers revealed that only 61 of the 5012 markers had 50 bp or more of their sequences in common within the 10 sorghum chromosomes (Supplementary Table S3). Furthermore, Fig. 1 shows that the previously reported SSR markers covered only a small proportion of sorghum's chromosomes. In particular, there was only one marker (Xtxp6) on the short arm of chromosome 6. Together, these results show that most of the SSR markers in the set designed here are unique and cover the regions that currently lack a sufficient number of SSR markers in each sorghum chromosome.

We mapped a set of 970 randomly chosen markers that had fewer AT/TA repeats to possible chromosome locations by means of ePCR and tested the results by means of fragment analysis. Forward and reverse primers were redesigned to provide accurate genotyping with post-PCR fluorescent labeling. AC, AG or AT was added to the 5' end of the forward primers, and GTTT was added to the 5' end of the reverse primers. PCR amplification was then performed and the PCR product was directly labeled with fluorescence-labeled R110-ddUTP by the single-tube method.²⁷ The labeled PCR products were then diluted and mixed with Hi-Di formamide containing a size standard. The heat-denatured products were then analyzed with an ABI Prism 3700 Genetic Analyzer, and the resulting allele data were analyzed with GeneMapper v3.7 software.

We screened for SSR polymorphisms among a total of 12 lines: four male sterility-maintenance sorghum lines (BTx623, BTx624, MS79B, Nakei MS-3B), seven inbred lines (74LH3213, Challwaxy Sorghum, JN43, SIL-05, bmr-6, Sennkinshiro, Takakibi) and one sudangrass line (Greenleaf). To do so, we used the following criteria to define the fragment size that would be used for genotyping. (i) If successful amplification was observed for BTx623 and the size difference was ≤ 5 bp between the fragment size closest to the expected PCR size in BTx623 and the expected fragment size in BTx623, then the corresponding SSR marker was used for the subsequent analysis. (ii) An SSR polymorphism was detected on the basis of a difference of ≥ 2 bp between the corresponding fragment sizes of each combination in the 12 lines. The mean frequency of total SSR polymorphisms between any two of the 12 lines (Table 2) was nearly identical to that used for calculating both the gene diversity index (D)²⁸ and the degree of differentiation (δ_T).²⁹

Of the 970 markers, 822 (84.7%) were identified as useful for the detection of polymorphisms; those excluded were susceptible to the experimental error peculiar to fragment analysis. These results are consistent with previous reports of the successful amplification of 90%⁸ and 85%⁷ of target SSRs. Moreover, 67.8% (658 of 970) of the SSR markers used for successful amplification in BTx623 revealed a PCR product size that corresponded to the expected product size (Supplementary Table S4). The genotypes of the 12 lines were determined by means of PCR amplification using the 658 primer sets. The mean value of the SSR polymorphisms was 45.1% between any two lines in all 12 lines, including sudangrass (Table 2). The combinations between pairs of sorghum lines (i.e. excluding sudangrass) showed a lower polymorphism rate (43.0%, Table 2). However, the mean difference was only 2.1%, which is not high. In previous studies, mean values of polymorphism rate for SSR loci were reported as 0.54⁶ and 0.69⁸, but these values were detected by using gel electrophoresis, which has lower sensitivity than fragment analysis. Moreover, in a previous report of fragment analysis,⁹ the estimated polymorphism rate (46.0%) was similar to our mean value of SSR polymorphisms (45.1%). This suggests that the SSR polymorphism rate might be affected more by differences in the materials used in the experiments than by the sensitivity of the analytical methods.

In general, the diversity of SSRs developed from cDNA-derived RFLP probe sequences is lower than the genetic variation detected by using SSR loci isolated from SSR-enriched genomic libraries. Schloss et al.⁹ also suggested that the low variation in SSR loci developed from RFLP probes was associated with a small proportion of dinucleotide motifs and differences in the length of the repeat units. Because of problems related to stutter and noise, we did not attempt to validate SSR markers with the $(AT/TA)_n$ motif by using fragment analysis. However, if a set of SSR marker motifs contains markers with an $(AT/TA)_n$ motif and an alternative highly sensitive method could be used for polymorphism detection, we would expect to see the same or a higher level of diversity than obtained here, since the AT-rich motif has a greater number of repeat units than most other motifs (Table 1).

We did not observe a clear correlation between the distributions of the number of SSR motifs on each chromosome. The distribution of the 558 SSR markers for which we detected polymorphisms in the combination of the 12 lines appears in Fig. 1. These markers were not remarkably localized on each chromosome compared with the distribution of 239 previously developed markers. These results suggest

Table 2. Mean rates (%) of SSR polymorphisms detected by means of fragment analysis in 11 sorghum lines and one sudangrass line (Greenleaf)

	BTx623	74LH3213	BTx624	Challwaxy Sorghum	JN43	MS79B	SIL-05	bmr-6	Nakei MS-3B	Sennkinshiro	Takakibi	Greenleaf
BTx623	41.9 (39.8)											
74LH3213	42.1 (40.4)	45.4 (43.0)					Mean ^a		Mean ^b			
BTx624	17.6 (17.6)	30.5 (28.9)	38.9 (36.8)									
Challwaxy Sorghum	32.7 (31.9)	40.1 (38.3)	30.2 (29.5)	41.5 (39.6)			45.1^c (42.7)		43.0^c (40.7)			
JN43	43.8 (40.3)	47.1 (43.7)	42.6 (39.1)	40.7 (38.0)	43.2 (40.1)							
MS79B	42.7 (41.0)	44.9 (43.0)	40.1 (38.4)	45.2 (43.7)	52.2 (49.4)	47.4 (45.3)						
SIL-05	47.3 (45.1)	49.3 (47.0)	43.0 (40.9)	38.1 (36.4)	45.8 (42.6)	47.9 (45.3)	46.1 (43.5)					
bmr-6	33.3 (31.5)	41.0 (39.1)	30.4 (28.6)	39.8 (38.4)	40.5 (38.5)	42.2 (41.4)	46.0 (43.9)	42.7 (40.6)				
Nakei MS-3B	36.5 (34.2)	41.9 (39.1)	30.2 (28.0)	29.2 (27.7)	42.6 (39.2)	40.7 (38.3)	40.2 (37.6)	33.8 (31.2)	41.6 (38.8)			
Sennkinshiro	52.4 (49.4)	51.7 (48.8)	53.2 (50.2)	51.6 (48.7)	25.0 (22.3)	53.3 (50.7)	48.3 (44.9)	52.0 (49.5)	53.7 (50.2)	48.1 (45.1)		
Takakibi	53.5 (50.8)	53.4 (50.8)	53.0 (50.3)	52.4 (49.5)	41.2 (38.0)	53.6 (51.3)	46.2 (43.1)	54.0 (51.8)	52.6 (49.4)	32.1 (30.2)	49.7 (46.9)	
Greenleaf	58.7 (55.5)	57.4 (54.3)	57.0 (53.8)	56.2 (53.1)	54.2 (50.5)	58.3 (55.2)	55.1 (51.9)	56.4 (53.2)	56.2 (52.2)	55.6 (51.5)	54.9 (50.8)	56.4 (52.9)

Values represent total frequencies of SSR polymorphisms. Values in parentheses are the frequencies of only polymorphisms with codominant markers between two lines. The mean value of each line is indicated by the diagonal boldfaced values.

^aTotal mean frequency of SSR polymorphisms between any two lines of the 12 lines, including the sudangrass line.

^bTotal mean frequency of SSR polymorphisms between any two of the 11 sorghum lines (i.e. excluding the sudangrass line).

^cSignificance is not observed between 45.1 and 43.0 at $P = 0.05$ levels.

that whether or not newly developed SSR markers include an (AT/TA)_n motif, these markers will be helpful in the construction of a linkage map for molecular dissection of sorghum's agronomic traits.

Validation data from our fragment analyses allowed estimation of the polymorphism frequency within the 5012 loci. When the rate of successful amplification with BTx623 was 67.8% and the mean value of the SSR polymorphisms was 45.1%, ~3400 (67.8% of 5012) were identified as suitable markers for the detection of SSR polymorphisms among *Sorghum* spp. lines and ~1500 (45.1% of 3400) markers could potentially indicate SSR polymorphisms in the combinations of *Sorghum* spp. lines. From this result and an estimated total genomic sequence length of 750 Mb, we estimated that the density of SSR markers developed here for use in polymorphism detection would be approximately one per 220 kb. However, because the gene-rich region and the newly developed markers tend to be located at positions away from the centromere, the distance between the SSR markers is likely to be <220 kb in the gene-rich regions. For this reason, the SSR markers developed here have the potential to be utilized not only as DNA markers for use in breeding but also as a means of map-based cloning. The use of public genomic sequences may provide additional markers in regions with fewer SSR markers.

Supplementary data: Supplementary data are available online at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported by a grant from the Ministry of Agriculture, Forestry and Fisheries of Japan (Genomics for Agricultural Innovation, SOR 0006).

References

1. Doggett, H. 1988, Sorghum, 2nd Ed., John Wiley and Sons, Inc.: New York.
2. Pereira, M. G., Lee, M., Bramel-Cox, P., Woodman, W., Doebley, J. and Whitkus, R. 1994, Construction of an RFLP map in sorghum and comparative mapping in maize, *Genome*, **37**, 236–243.
3. Nagamura, Y., Tanaka, T., Nozawa, H., Kaidai, H., Kasuga, S. and Sasaki, T. 1998, Syntenic regions between rice and sorghum genomes, In: Nakagawa, H., and Kobayashi, M. (eds.), *Proceedings of the International Workshop; Utilization of Transgenic Plant and Genome Analysis in Forage Crops*, National Grassland Research Institute: Tochigi, pp. 97–103.
4. Bowers, J. E., Abbey, C., Anderson, S., 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.
5. Menz, M. A., Klein, R. R., Mullet, J. E., Obert, J. A., Unruh, N. C. and Klein, P. E. 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.
6. Brown, S. M., Hopkins, M. S., Mitchell, S. E., et al. 1996, Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench], *Theor. Appl. Genet.*, **93**, 190–198.
7. Bhatramakki, D., Dong, J., Chhabra, A. K. and Hart, G. E. 2000, An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench, *Genome*, **43**, 988–1002.
8. Kong, L., Dong, J. and Hart, G. E. 2000, Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs), *Theor. Appl. Genet.*, **101**, 438–448.
9. Schloss, J., Mitchell, E., White, M., et al. 2002, Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench, *Theor. Appl. Genet.*, **105**, 912–920.
10. Tao, Y. Z., Jordan, D. R., McIntyre, C. L. and Henzell, R. G. 1998, Construction of a genetic map in a sorghum recombinant inbred line using probes from different sources and its comparison with other sorghum maps, *Aust. J. Agric. Res.*, **49**, 729–736.
11. Taramino, G., Tarchini, R., Ferrario, S., Lee, M. and Pé, M. E. 1997, Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*, *Theor. Appl. Genet.*, **95**, 66–72.
12. Arumuganathan, K. and Earle, E. D. 1991, Nuclear DNA content of some important plant species, *Plant Mol. Biol. Rep.*, **9**, 208–218.
13. Doebley, J., Durbin, M., Golenberg, E. M., Clegg, M. T. and Ma, D. P. 1990, Evolutionary analysis of the large subunit of carboxylase (rbcL) nucleotide sequence among the grasses (Gramineae), *Evolution*, **44**, 1097–1108.
14. Sobral, B. W. S., Braga, D. P. V., LaHood, E. S. and Keim, P. 1994, Phylogenetic analysis of chloroplast restriction enzyme site mutations in the *Saccharinae* Griseb. subtribe of the *Andropogoneae* Dumort. tribe, *Theor. Appl. Genet.*, **87**, 843–853.
15. Bedell, J. A., Budiman, M. A., Nunberg, A., et al. 2005, Sorghum genome sequencing by methylation filtration, *PLoS Biol.*, **3**, e13.
16. Paterson, A. H., Bowers, J. E., Bruggmann, R., et al. 2009, The *Sorghum bicolor* genome and the diversification of grasses, *Nature*, **457**, 551–556.
17. Schuler, G. D. 1997, Sequence mapping by electronic PCR, *Genome Res.*, **7**, 541–550.
18. Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S. and McCouch, S. 2001, Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential, *Genome Res.*, **11**, 1441–1452.
19. Fukuoka, H., Nunome, T., Minamiyama, Y., Kono, I., Namiki, N. and Kojima, A. 2005, Read2Marker: a data

- processing tool for microsatellite marker development from a large data set, *Biotechniques*, **39**, 472, 474, 476.
20. Rozen, S. and Skaletsky, H. 2000, Primer3 on the WWW for general users and for biologist programmers, In: Misener, S., and Krawetz, S. A. (eds.), *Methods in Molecular Biology*, Humana Press Inc: Totowa, pp. 365–386.
21. McCouch, S. R., Teytelman, L., Xu, Y., et al. 2002, Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.), *DNA Res.*, **9**, 199–207.
22. Yu, J. K., Dake, T. M., Singh, S., et al. 2004, Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat, *Genome*, **47**, 805–818.
23. Thiel, T., Michalek, W., Varshney, R. and 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.
24. Cordeiro, G. M., Casu, R., McIntyre, C. L., Manners, J. M. and Henry, R. J. 2001, Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum, *Plant Sci.*, **160**, 1115–1123.
25. Asp, T., Frei, U. K., Didion, T., Nielsen, K. K. and Lubberstedt, T. 2007, Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*, *BMC Plant Biol.*, **7**, 36.
26. Kantety, R. V., La Rota, M., Matthews, D. E. and Sorrells, M. E. 2002, Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat, *Plant Mol. Biol.*, **48**, 501–510.
27. Inazuka, M., Tahira, T. and Hayashi, K. 1996, One-tube post-PCR fluorescent labeling of DNA fragments, *Genome Res.*, **6**, 551–557.
28. Nei, M. 1973, Analysis of gene diversity in subdivided populations, *Proc. Natl. Acad. Sci. USA*, **70**, 3321–3323.
29. Gregorius, H. R. 1987, The relationship between the concepts of genetic diversity and differentiation, *Theor. Appl. Genet.*, **74**, 397–401.