Theor Appl Genet (1996) 93: 190–198

S. M. Brown · M. S. Hopkins · S. E. Mitchell
M. L. Senior · T. Y. Wang · R. R. Duncan
F. Gonzalez-Candelas · S. Kresovich

Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]

Received: 3 October 1995 / Accepted: 19 January 1996

Abstract Simple sequence repeats (SSRs), also known as microsatellites, are highly variable DNA sequences that can be used as markers for the genetic analysis of plants. Three approaches were followed for the development of PCR primers for the amplification of DNA fragments containing SSRs from sorghum [*Sorghum bicolor* (L.) Moench]: a search for sorghum SSRs in public DNA databases; the use of SSR-specific primers developed in the Poaceae species maize (*Zea mays* L.) and seashore paspalum grass (*Paspalum vaginatum* Swartz); and the screening of sorghum genomic libraries by hybridization with SSR oligonucleotides. A total of 49 sorghum SSR-specific PCR primer pairs (two designed from GenBank SSR-containing sequences and 47 from the sequences of genomic clones) were screened on a panel of 17 sorghum and one maize accession. Ten primer pairs from paspalum and 90 from maize were also screened for polymorphism in sorghum. Length polymorphisms among amplification products were detected with 15 of these primer pairs, yielding diversity values ranging from 0.2 to 0.8 with an average diversity of 0.56. These primer pairs are now available for use as markers in crop improvement and conservation efforts.

Key words Microsatellite · Germplasm · Genetic resources · Genetic analysis

Communicated by G. E. Hart

S. M. Brown · M. S. Hopkins · S. E. Mitchell · T. Y. Wang
S. Kresovich (✉)
USDA-ARS, Plant Genetic Resources Conservation Unit,
Griffin, GA 30223-1797, USA

R. R. Duncan
Crop and Soil Sciences Department, University of Georgia,
Griffin, GA 30223-1797, USA

M. L. Senior
USDA-ARS, Molecular Marker Facility,
Department of Crop Science, North Carolina State University,
Raleigh, NC 27695-7620, USA

F. Gonzalez-Candelas
Department of Genetics, University of Valencia,
Valencia 46100, Spain

Introduction

The use of molecular markers has become widely accepted as a valuable tool for plant-breeding programs as well as for evolutionary and conservation studies. Simple sequence repeats (SSRs), also known as microsatellites, are a relatively new class of molecular markers based on tandem repeats of short (2–6 bp) DNA sequences (Litt and Lutty 1989). These repeats are highly polymorphic, even among closely related cultivars, due to mutations causing variation in the number of repeating units. Variations in tandem repeat length accumulate in populations more rapidly than the point mutations and larger insertion/deletion events responsible for most restriction fragment length polymorphisms. Different alleles can be detected at a locus by the polymerase chain reaction (PCR), using conserved DNA sequences flanking the SSR as primers.

SSRs offer a potentially attractive combination of features that are useful as molecular markers. First, SSRs have been reported to be highly polymorphic and thus highly informative in plants, providing many different alleles for each marker screened, even among closely related individuals (Akkaya et al. 1992; Saghai- Maroof et al. 1994). Second, SSRs can be analyzed by a rapid, technically simple, and inexpensive PCR-based assay that requires only small quantities of DNA. Third, SSRs are co-dominant and simple Mendelian segregation has been observed. Finally, SSRs are both abundant and uniformly dispersed in both human (Weber 1990) and plant genomes (Lagercrantz et al. 1993; Wang et al. 1994, Akkaya et al. 1995). The primary disadvantage of SSRs as molecular markers is the cost and research effort required to clone and sequence SSR-containing DNA fragments from the plant species of interest.

The goal of the present project was to identify highly polymorphic SSRs in sorghum [*Sorghum bicolor* (L.)

Moench] and to create PCR primer pairs that could be used to determine genotypes of large numbers of germplasm accessions. Molecular characterization could help curators to identify redundancies and gaps in sorghum collections. These data would also allow the clustering of accessions by a neutral genotypic measure, which could facilitate the creation of core subsets from large collections and would be highly valuable to breeders and other users of the collection.

To develop sorghum SSR markers, a strategy of constructing genomic libraries, screening by hybridization with SSR probes, and sequencing positively hybridizing clones was followed (Condit and Hubbell 1991; Thomas and Scott 1993; Lavi et al. 1994). Because library screening and sequencing are laborious and costly, two additional methods to identify SSRs were investigated. First, a search was conducted for SSRs in public sequence databases. Second, as suggested by recent work using maize (*Zea mays* L.) RFLP probes in sorghum (Hulbert et al. 1990; Whitkus et al. 1992; Pereira et al. 1994), primer pairs designed from sequence information in maize and seashore paspalum (*Paspalum vaginatum* Sw.) were tested for the ability to amplify SSRs in sorghum. Several investigators have demonstrated the amplification of SSRs across genera of mammals (Moore et al. 1991; Ammer et al. 1992; Kondo et al. 1993) and plants (Kresovich et al. 1995) using conserved primers. This three-pronged approach of database search, inter-species homology, and the screening of genomic libraries could be applied to the development of SSR markers for any plant species. Based on the results of our efforts to develop SSR markers for sorghum, an analysis of the relative merits of these three approaches is provided.

Materials and methods

Plant material

Seventeen sorghum lines and maize breeding line B73 were used in this study as the primary screening array to measure the diversity of sorghum SSR markers. The sorghum lines included temperately and tropically adapted material representing B lines (male sterility maintainers): Martin, B97E, B68181, BSD106, BT×8618; and R lines (male fertility restorers): 38 M, 58 M, RT×430, RTAM428, GC103, GC104; as well as converted accessions originally from Africa and India: SC326 (IS 3758), SC414 (IS 2508), SC574 (IS 8337), SC599 (IS 17459), SC689 (IS 2729), SC971 (Stephens et al. 1967). All are photoperiod-insensitive genotypes. A second array of four genetically dissimilar sorghum lines, B35 (Scheuring and Miller, 1978), Tx7000 (Karper 1949), CK60 (Karper 1953), and PI229828 (Pereira et al. 1994), were screened with the sorghum SSR markers to determine if additional polymorphisms could be found in a more diverse spectrum of sorghum germplasm.

Database screening of sorghum sequences for SSRs

A search was performed using the FINDPATTERNS program in the GCG package (Wisconsin Sequence Analysis Package, Version 7.3, Genetics Computer Group, Madison, Wis. 1994) on all sorghum sequences included in GenBank and EMBL updated until 7/27/94 [GenBank release 83.0 (6/94), GenNew (to 7/25/94), and EMBL (modified) release 38.0 (3.94)] for all possible di-, tri-, and tetra-nucleotide repeats with six or more repeat units. Primer pairs were derived from these sequences using the DesignerPCR™ program (Research Genetics) and synthesized on an Applied Biosystems Inc. 392 nucleic-acids synthesizer.

DNA extraction

Plants were grown from seed for 3–4 days in the dark, then etiolated hypocotyls were harvested and lyophilized. Dry tissue was ground into a fine powder, using a mortar and pestle and liquid nitrogen. DNA was extracted using the CTAB method of Saghai-Maroof et al. (1984), then further purified by an RNase digestion followed by extraction with phenol/chloroform/iso-amyl alcohol (25:24:1) and ethanol precipitation.

Library construction and screening

A size-selected sorghum genomic library was constructed in the vector LambdaZap II™ (Stratagene). Genomic DNA from cultivar RT×430 was digested with *Eco*RI. Fragments sized 200–600 bp were isolated from low-melting agarose (Sea Plaque™, FMC) with the Prep-A-Gene™ kit (Bio-Rad) and ligated into the *Eco*RI site of LambdaZap II™, followed by packaging with the Gigapack Gold™ packaging extract and transfection of XL1-Blue cells (Stratagene). Each ligation/packaging reaction produced approximately 150 000 recombinant clones.

The phage library was screened by plaque hybridization with a mixture of radiolabeled $(GA)_{10}$ and $(CA)_{10}$, $(GAA)_7$, $(ATT)_7$, and $(ATTT)_5$ oligonucleotides. Probes were prepared by random hexamer labeling with ^{32}P-dATP (Feinberg and Vogelstein 1983) using the Random Primers DNA Labeling Kit (GIBCO/BRL). For screening, phage clones (2000 plaques/150-mm plate) were lifted onto nylon filters (Magna Lift™, MSI Inc.), denatured, and fixed by baking at 80°C for 1 h. Filters were hybridized in 6×SSPE, 1% SDS at 55°C; washed twice in 2×SSPE, 0.1% SDS at 55°C; washed once in 1×SSPE, 0.1% SDS at 55°C, then autoradiographed.

Positive clones identified by the plaque hybridization assay were converted to pBluescript™ plasmids by in vivo excision (Stratagene protocol). A set of four plasmid clones derived from each positively hybridizing plaque were then screened by colony hybridization, using the same probe mixture as in the plaque hybridization, to confirm the presence of SSRs. Plasmids with SSR-containing inserts were sequenced on the Applied Biosystems Inc. 373 automated sequencer using the dye terminator cycle sequencing kit (Perkin Elmer).

PCR analysis

Sorghum and seashore paspalum PCR primer pairs were designed from sequence flanking SSRs using the Designer PCR program (Research Genetics) and synthesized on an Applied Biosystems Inc. 392 nucleic-acids synthesizer. Pairs of primers (length 15–30 bp) were designed with the annealing temperature (T_m) limited to 55–65°C, the T_m variance limited to 3°C, and the product length in the range of 85–400 bp. Maize SSR-specific PCR primer pairs (from SSR sequences identified in GenBank) were designed using the program Primer 0.5 (provided by S. Lincoln, M. Daly and E. Lander, Cambridge, Mass.). Seashore paspalum SSR-specific PCR primer pairs were provided by Z. Liu, USDA-ARS, Griffin, Ga. (Liu et al. 1995).

All PCR reactions were performed on a Perkin Elmer 9600 thermocycler in 25-µl volumes containing 50 ng of sorghum genomic DNA, 10 pmol (50 ng) of each primer, 2 mM $MgCl_2$, 40 µM each dNTP, 1× PCR buffer [50 mM KCl, 10 mM Tris-HCl pH (8.0), 10 mM NaCl, 0.01 mM EDTA, 0.5 mM DTT, 0.1% Triton X-100], and 0.5 units of *AmpliTaq*™ DNA polymerase (Perkin Elmer).

Each pair of primers was initially screened for amplification of a specific product from sorghum genomic DNA, using a 65–55°C "touchdown" PCR technique (Mellersh and Sampson 1993). The annealing temperature for the initial 18 cycles was decreased by one

degree every two cycles from 65°C to 56°C, followed by 20 cycles at 55°C [94°C-4 min initial denaturation; 94°C-30 s, 65–56°C-30 s, 72°C-1 min, for 2 cycles at each temperature; then 94°C-15 s, 55°C-30 s, 72°C-1 min for 20 cycles; and 72°C-10 min final extension]. DNA fragments produced by these amplifications were separated by electrophoresis in 3% Metaphor™ (FMC BioProducts) agarose gels and stained with ethidium bromide.

In order to obtain more precise estimates of fragment sizes and to identify small size differences between fragments, polymorphic PCR products were electrophoresed on 8% non-denaturing polyacrylamide gels stained with ethidium bromide. A simplified PCR protocol with the number of cycles reduced to 25 and a constant 55°C annealing temperature [5 min-95°C initial denaturation, (95°C-30 s, 55°C-45 s, 72°C-60 s) for 25 cycles, and a 72°C-10 min final extension] was used for amplifications that were separated on acrylamide.

All fragment sizes were estimated based on linear interpolation from the GIBCO/BRL 1-kb size marker. The level of polymorphism of each marker was estimated with a diversity index D [$D=1-\sum p_i^2$ in which p_i is the frequency of the i^{th} SSR allele (Saghai Maroof et al.1994)].

Results

Database search

The database search resulted in two positive matches, corresponding to locus SBKAFGK1 (accession # X16104), a kafirin gene, and locus SVHRGPG (accession # X56010), a hydroxyproline-rich glycoprotein gene. As reported in the database, SBKAFGK1 contains an $(AAC)_{10}$ repeat and SVHRGPG contains a $(CGC)_9$. Both database entries had sufficient flanking sequences to allow primer design. The primer pair flanking the SBKAFGK1 SSR locus amplified a fragment that exhibited length polymorphisms across a test array of 17 sorghum lines. Primers for SVHRGPG amplified a fragment that was monomorphic in the test array.

Use of maize primers to amplify sorghum SSRs

Twenty three primer pairs generated from SSRs found in maize GenBank sequences were screened on sorghum lines (Table 1) using touchdown PCR (65–55°C). These amplifications yielded a variety of results: (1) no products, (2) multiple products ranging from 20 to >1000 bp [similar to the pattern of bands produced from single primer RAPD (random amplified polymorphic DNA) amplifications], (3) a discrete monomorphic band, or (4) a discrete polymorphic band. Only one primer pair (ZM-ADH2N) produced discrete, reproducible bands that were polymorphic in the test array of 17 sorghum lines (Figs. 1,2). Southern blots of these fragments, using an SSR oligonucleotide as a probe, verified that the polymorphic products produced by primer pair ADH2N and from some of the monomorphic primer pairs contained SSRs (data not shown).

In parallel work, an additional set of 67 maize SSR-specific primer pairs derived from GenBank sequences was tested on a panel of four genetically dissimilar sorghum lines (B35, Tx7000, CK60, and PI229828). All had been previously determined to be polymorphic in maize (M.L. Senior, unpublished data). Forty-three of the amplifications yielded a product from sorghum similar in size to that in maize; 12 amplifications showed polymorphism among the four sorghum lines.

Use of seashore paspalum primers to amplify sorghum SSRs

A total of ten primer pairs, previously identified as polymorphic in seashore paspalum (Liu et al.1995), were screened on a panel of 17 sorghum lines. None of these primer pairs yielded discrete products in sorghum similar in size to those observed in seashore paspalum. Multiple band patterns (similar to the products from RAPD reactions) were produced and no clear polymorphisms were detected among lines.

Library screening

Approximately 120000 lambda clones were screened by plaque hybridization using a combination of $(GA)_{10}$, $(CA)_{10}$, $(GAA)_7$, $(ATT)_7$, and $(ATTT)_5$ oligonucleotides as radiolabeled probes. Only 3–5 positives per filter (approximately 2000 plaques/filter) were found in the initial screening, an efficiency of approximately 0.2%. A second screening of these positive clones by hybridization to excised inserts (in pBluescript plasmids) eliminated approximately 70% as false positives. The remaining clones were sequenced. Approximately 30% of sequenced clones were suitable for the design of primer pairs flanking an SSR. Clones were rejected as unusable due to insufficient flanking sequence, repeats that contained too few tandem copies (<14 bp), or severely imperfect repeats. A total of 146 clones were isolated and sequenced from the sorghum genomic library, yet only 47 of these sequences were suitable for the synthesis of primer pairs flanking SSRs.

Similarities between the sequences of sorghum SSR clones isolated from the library and entries in the GenBank database were determined by the BLAST local similarity search program (Altschul et al. 1990). Only regions flanking the SSR were used for BLAST searching to avoid spurious homology with other SSRs. Surprisingly, none of the database entries homologous to these flanking regions contained SSRs themselves. The most significant similarities were found with sorghum, maize, and rice nuclear genes, cDNAs and chloroplast sequences. However, several highly significant homologies to human and other sequences were detected (Table 2). In all cases where the homology between SSR-containing sorghum clones and known genes was strong enough to permit an alignment, the SSRs were found to be located outside of coding regions.

Fifteen of the forty seven primer pairs (32%) amplified putative polymorphic loci in the test array of 17 sorghum lines; 21 pairs (45%) were monomorphic; and 11 pairs (23%) produced either no amplification product or multiple bands. All monomorphic and polymorphic PCR products were found to have similar sizes to that predicted from

Table 1 Sorghum DNA amplified with maize SSR-specific primers[a]

Primer ID#	Primer sequence[b]	Product size	Maize size	Notes
MZE-BZIP	F:ACTGTTCCACCAAACCAAGC R:TCCTCCTACTCCACCTGGTG	175	150	Monomorphic
ZM-GPC1	F:CCTCTACTCGCCAGTCGC R:TTTGGTCAGATTTGAGCACG	None	148	
ZM-ADH2N	F:TGCGAAGAAAGCAGTAGCAAA R:TGGAGGTAGAAGACGCACG	200	149	Polymorphic
MZE-TPI1	F:CCCCTTGTCTTTCTTCCTCC R:CGATTAGATTGGGGTGCG	None	115	
MZE-TRACTB (CT)	F:CGAAAGTCGATCGAGAGACC R:TTTATTTACCGGCTTTCTGGG	None	120–130	
MZE-TRACTB (ACGT)	F:ACATGAGCTGACGACGAGC R:ATTTCATTGTCTGTTCGGCC	None	150	
MZE-PPDKA2	F:TAATTTAAACACCACACCACCG R:ACACACGCCAAAGAAAAACC	500, 200–300	100–200	Polymorphic and Monomorphic bands
ZM-GLB10	F:AGGAACCCGGAATCGTTC R:CAACGCGATCGATGTGAG	None	142	
ZM-GPA1	F:CTACCACCGTGCCACTCC R:GCTAGAGTGCAAGCCAAACC	500	108	Monomorphic
MZE-ZEINP	F:GGATGATGGACGTGCAGTC R:CTGGTACTGGTAGAGTCCACCC	None	150	
MZE-GSTIA1	F:CCCTCTCCCTTTCCCTTTC R:GAAACAAAACCCATCGCG	None	100–200	
MZE-WXBA	F:TGTGGAAGTGGAGCCCAG R:TGCTAGCTGCTGCTGGTG	None	130–150	
MZE-PPR	F:CCTGTTCGACATGGACTGG R:AGCCGTCAGAACGAGTCG	None	120–130	
MZE-CAT3GN (AT)	F:TCGCCAGGAACAAGAGAAAT R:ATTCTCCCGCATAGATGTGC	None	100	
MZE-CAT3GN (ATT) (nc020)	F:CATGAAGGCAAACATGTGATG R:TCGATACACATCCATCTTCAGG	None	110–120	
MZE-CAT3GN (ATTT)	F:AGCGACTTTGACGTTTCGAT R:CGCGCAGTTCTTTCTTGTGA	250	100	Monomorphic
MZE-MTMSODC	F:TACGCACCTACGAGAGTCGA R:CGAAGTCGTAGGAGAGGTCG	None	170–180	
MZE-PLBH	F:AGGAGGATCTCATCGTCCG R:CATTATCTGTTCGGCCCG	200	150	Monomorphic
MZE-TNENSPM	F:GGATGATTTTTTCATAGGCTCG R:CCGTCACCAATTGCTAACCT	None	140	
ZM-27KZNA	F:TCTTTATAATCAACCCGCACTC R:TATATGCCCGTGCATTGCTA	None	130	
ZM-MAYSPG	F:CCTGTTCGACATGGACTGG R:AGCCGTCAGAACGAGTCG	250	175	Monomorphic
ZM-GST3	F:TTACTCCTATCCACTGCGGC R:GTCGAGGCCCTTCTCGTT	None	130–150	
ZM-ZC2ZEI	F:GACCTTGGTTGTTGGCTCAT R:CCCTTTCGATCCAGACGTAA	450	200	Monomorphic
ZM-ZEIN27	F:TGTTGGCTCCATATTCACACA R:TATATGCCCGTGCATTGCTA	550	80–100	Monomorphic

[a] All PCR amplifications were conducted using 65–55°C "touchdown" cycle conditions (Mellersh and Sampson1993).
[b] From Senior and Huen 1993; Senior et al. 1995, and unpublished data from M. L. Senior

the sequence of the original cloned genomic fragment (Table 3). For each of the polymorphic primers, 2–6 different fragment sizes were detected in amplifications from the array of 17 sorghum lines. When these same 47 primer pairs were screened on a set of four genetically dissimilar sorghum lines (which were used previously as parents to generate highly heterozygous F_2 mapping populations), polymorphisms were found with an additional six of the previously monomorphic primer pairs.

Six primer pairs that initially produced no product were re-designed based on additional flanking sequences. Two of these re-designed primers produced monomorphic products and two produced multiple bands, but no additional polymorphic SSRs were identified.

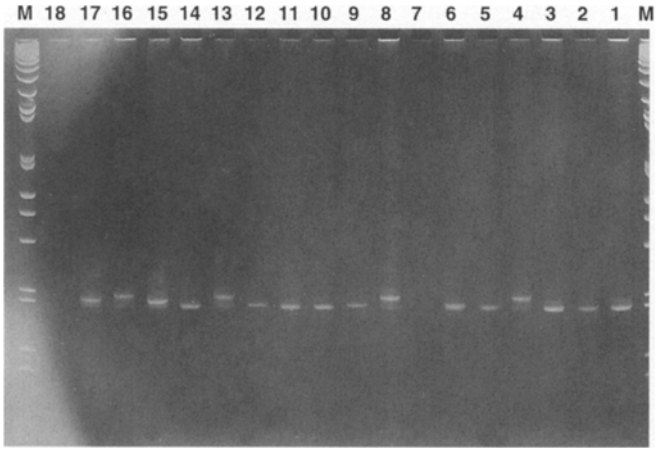

Fig. 1 Photograph of EtBr-stained polyacrylamide gel of polymorphic PCR products from 17 sorghum lines (*lanes: 1*=RT×430, *2*=58 M, *3*=BT×8618, *4*=Martin, *5*=38 M, *6*=B68181, *7*=SC414, *8*=B97E, *9*=SC326, *10*=RTAM428, *11*=SC599, *12*=GC104, *13*=BSD106, *14*=GC103, *15*=SC689, *16*=SC574, *17*=SC971) and maize line B73 (*lane 18*) amplified with SSR-specific sorghum primer pair Sb5-085. First and last lanes (labeled *M*) contain the 1-kb size marker from GIBCO/BRL

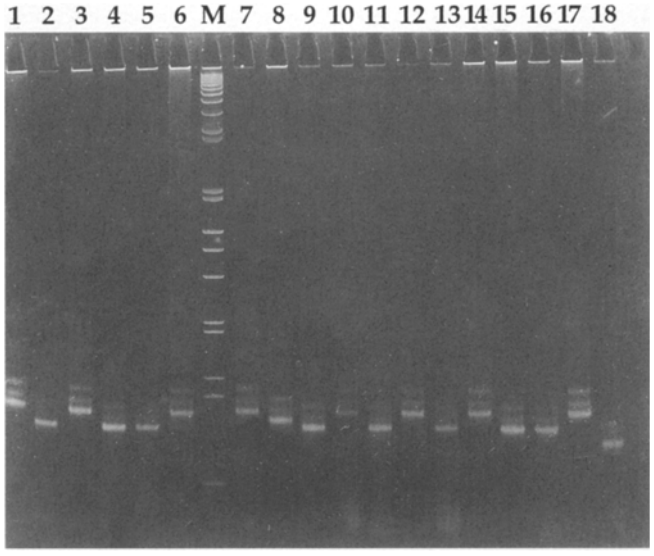

Fig. 2 Photograph of EtBr stained polyacrylamide gel of polymorphic PCR products from 17 sorghum lines (*lanes: 1*=SC971, *2*=SC574, *3*=SC689, *4*=GC103, *5*=BSD106, *6*=GC104, *7*=SC599, *8*=RTAM428, *9*=SC326, *10*=B97E, *11*=SC414, *12*=B68181, *13*=38 M, *14*=Martin *15*=BT×8618, *16*=58 M, *17*=RT×430) and maize line B73 (*lane 18*) amplified with SSR-specific primer pair ZM-ADH2N that was derived from a maize GenBank accession. The lane labeled M (between lanes 6 and 7) contains the 1 kb size marker from GIBCO/BRL

Table 2 BLAST[a] similarity results for SSR clones

Clone	Similar database entry	Significance of similarity (*P* value)	Comments
Sb1-13	Maize chloroplast	*P*=1.9e-32	Non-coding region
Sb5-214	Sorghum opaque-2 gene	*P*=7.0e-31	In 1st intron does not include SSR
SB4-034	Rice cDNA S4208_1A *Arabidopsis* cDNA 163D2T7	*P*=1.3e-10 *P*=1.6e-4	
Sb1-001	*Arabidopsis* cDNA VCVCE06	*P*=3.3e-10	
Sb5-236	*C. lacryma-jobi* opaque-2 gene Sorghum pepC gene	*P*=1.3e-7 *P*=0.003	
Sb5-30	Human cDNA 83898	*P*=1.1e-6	Similar to human fatty acid binding protein
Sb4-032	Human EST 21ES0035	*P*=1.1e-4	
Sb6-306	*C. coturnix* clusterin gene	*P*=2.8e-4	
Sb6-101	Human cDNA 70854	*P*=9.6e-3	Similar to human 60S ribosomal protein L37A

[a] Basic local alignment search tool (Altschul et al. 1990)

Discussion

Analysis of the three methods

The efficacy of three methods of obtaining SSR markers in sorghum was investigated: searching databases; PCR amplifications with SSR-specific primers designed for other Poaceae species; and screening genomic libraries by hybridization and sequencing of clones. Generalizations based on these comparisons between the three methods are subject to the constraints of small sample size, particularly with respect to database searching, but these findings provide a practical example of three approaches to SSR marker development.

Table 3 Sorghum SSR primers. P=Polymorphic, M=Monomorphic

SSR ID	Primers	$T_{opt.}$	Repeat	Product Size[a]	Type of product	# of frags.	Diversity index[b]
SBKAFGK1	F:GCTTTCGGCGAGCATCTTACAA R:GCGGTTGGATTCGCCATG	60	$(AAC)_9$	280–320	P	3	0.48
SVHRGPG	F:CAAGCCGCCCACCTCAACT R:ATCGATCACCTTCTGCAACACAAA	61	$(CGG)_8$	150	M	1	
ZMADH2N	F:TGCGAAGAAGCAGTAGCAAA R:TGGAGGTAGAAGACGCACG	60	$(AG)_7$	110–120	P	3	0.56
Sb1-1	F:TCCTGTTTGACAAGCGCTTATA R:AAACATCATACGAGCTCATCAATG	60	$(AG)_{16}$	260–300	P	3	0.47
Sb1-10	F:GTGCCGCTTTGCTCGCA R:TGCTATGTTGTTTGCTTCTCCCTTCTC	65	$(AG)_{27}$	350–400	P	4	0.51
Sb1-11	F:CGCAAGGAGAACGCAGAATCC R:TGATGCCGTCCAACCCTGAG	60	$(AG)_{14}$	275	M	1	
Sb1-12	F:AGACCAATCCAGCAATGAGTCC R:AAAATGTTAGGGAGGAGAGGTTGAC	58	$(AG)_{29}$	200	M	1	
Sb1-11	F:GCCTGTATTGGGCTTAGTCT R:CCGCAGCAGCAGAAT	57	$(AG)_{14}$	270	M	1	
Sb4-7	F:ATTGCTGGAATGGTCAGTC R:TGCCCCAAATGTAGGAC	57	$(AG)_{15}$	247	M	1	
Sb4-15	F:GCTGCTAAGCCGTGCTGA R:TTATTTGGGTGAAGTAGAGGTGAACA	57	$(AG)_{16}$	120–130	P	4	0.62
Sb4-22	F:TGAGCCGAAAACCGTGAG R:CCCAAAACCAAGAGGGAAGG	59	$(ACGAC)_4$/ $(AG)_6$	270–300	P	5	0.66
Sb4-32	F:CTCGGCGGTTAGCACAGTCAC R:GCCCATAGACAGACAGCAAAGCC	59	$(AG)_{15}$	160–180	P	6	0.70
Sb4-51	F:ATCCCCTACCAAATGTATCCTAAATC R:GGGCGACCAAAGGTGTATG	59	$(AG)_{12}$	350	M	1	
Sb-4-72	F:TGCCACCACTCTGGAAAAGGCTA R:CTGAGGACTGCCCCAAATGTAGG	60	$(AG)_{16}$	190	M	1	
Sb4-121	F:GAAAAATCTCCGTCAATCCCAAAATAA R:CGCTGAACAACGAAAGGAATAAGTG	60	$(AC)_{14}$	200–225	P	6	0.45
Sb4-141	F:GCAGCAGGAATACACACACGAGG R:TACCGCACTACACTACACGAAATCATCT	60	$(CAG)_{12}$ imperfect	375	M	1	
Sb5-30	F:CGATAACTCTCTTTCCGATTTT R:CTCACCTCGTCATTCTTTTCT	57	$(AAG)_{27}$/ $(AGG)_{10}$	170	M	1	
Sb5-85	F:AGACGCTTTTCTCTCTCTCTCTCTCTCT R:TAGCCCTGCCGCATACTGAATG	60	$(AG)_{12}$	200–225	P	4	0.80
Sb5-206	F:ATTCATCATCCTCATCCTCGTAGAA R:AAAAACCAACCCGACCCACTC	57	$(AC)_{13}$/ $(AG)_{20}$	115–130	P	3	0.21
Sb5-208	F2:CTTTGATGCGAAGGCAGT R2:ACTTTTTCCTTTGTCTTCCCCT	59	$(AG)_{19}$	292	M	1	
Sb5-210	F:TCAAGGAAAACCGTTCAA R:ACCTCCACTCTATCTCTCTCAAT	57	$(AG)_8$	241	M	1	
Sb5-214	F:CGCCGCACCCGTCTA R:CGGAGGGGGAACACG	59	$(AG)_{14}$	174–300	P	2	0.22
Sb5-226	F:TCCAGAGGTAGTGGTTTGATAA R:AAGACACACGATTACACGAAGAGT	55	$(AG)_9$	120	M	1	
Sb5-236	F:GCCAAGAGAAACACAAACAA R:AGCAATGTATTTAGGCAACACA	57	$(AG)_{20}$	165–185	P	3	0.68
Sb5-256	F2:AATTTGCTTTTTGGTCCGTTT R2:TAGGAAAGACAGTACTAGAGGTCA	58	$(AG)_8$	169	M	1	
Sb5-258	F:TAATCACCTTTGAATCTCCATCTC R:GGGTGGCCACCGAAGAGT	58	$(AG)_{24}$	180	M	1	
Sb6-34	F:AACAGCAGTAATGCCACAC R:TGACTTGGTAGAGAACTTGTCTTC	60	$(AC)/(CG)_{15}$	190	M	1	
Sb6-36	F:GCATAATGACGGCGTGCTC R:CTTCCAAGTGAAAGAAACCATCA	60	$(AG)_{19}$	155–190	P	5	0.73
Sb6-42	F:TTTTCCTCTTTCAGATAACCGTA R:CCCACCAAGGGCATC	57	$(AG)_{26}$	180	M	1	
Sb6-48	F:AATTCGATCAACTACTTTTTACCA R:TGCGTTGGGCTCTGC	56	$(AG)_{18}$	136	M	1	

Table 3 Continued

SSR ID	Primers	$T_{opt.}$	Repeat	Product Size[a]	Type of product	# of frags.	Diversity index[b]
Sb6-57	F:ACAGGGCTTTAGGGAAATCG R:CCATCACCGTCGGCATCT	60	$(AG)_{18}$	285–305	P	4	0.46
Sb6-84	F:CGCTCTCGGGATGAATGA R:TAACGGACCACTAACAAATGATT	58	$(AG)_{14}$	170–190	P	5	0.69
Sb6-306	F:CACGTCGAAATTAACCCTCACTAAA R:GGAGCGAGCAAGCCTACTG	56	$(TGGA)_6$	(253)	M	1	
Sb6-313	F:TTCTTCAGTTCGCACAGCATAA R:ACCTGCAGTGCACTTGACTATTG	54	$(AC)_{10}$	128	M	1	
Sb6-325	F:AGCGCAGGAGCGCGAA R:TCATCCGCTACTACCGTCAGAAA	56	$(AAG)_{22}$	110–140	P	2	0.60
Sb6-327	F:TCTCTACAACCAAATTAATAGGTGGGTC R:TCCTTGGACAATACCCCTTACAC	56	$(CTT)_9$	154	M	1	
Sb6-342	F:TGCTTGTGAGAGTGCCTCCCT R:GTGAACCTGCTGCTTTAGTCGATG	56	$(AC)_{25}$	250–320	P	3	0.31
Sb6-350	F:TCAGGCAACGGATAACTCATACT R:TCCTGCTCCGGCAGTCGAA	56	$(AC)_{17}$ imperfect	290	M	1	

[a] Fragment sizes were estimated based on linear interpoloation from the GIBCO/BRL 1-kB size markers run in the first and last lanes of two replicate 8% polyacrylamide gels

[b] Diversity indices were calculated with the formula $1-\Sigma p_i^2$, in which p_i is the frequency of the i^{th} SSR allele (Saghai Maroof et al 1994)

Database screening was the least costly in terms of time and resources and has been used as an entry point for the development of SSR markers in several plant species including soybean (*Glycine max* L.) (Akkaya et al. 1992), maize (Senior and Heun 1993), *Arabidopsis thaliana* (Bell and Ecker 1994), and barley (*Hordeum vulgare* L.) (Saghai-Maroof et al. 1994). However, sorghum is a relatively poorly characterized species, with only 45 entries available in GenBank (release 87.0, 1995); thus it was not surprising that a search yielded only two SSRs. Since one of these two loci was found to be polymorphic, the use of less-stringent search criteria (including repeats of fewer than 16 bp and imperfect repeats) and additional search strings (penta-, and hexa-meric repeat motifs) might be worthwhile in other searches for SSRs in the database.

Investigators working on other species should have success with database searches in direct proportion to the amount of sequence information that they can access. Only in a few cases, such as *Arabidopsis* (15 010 entries), rice (*Oryza sativa* L.) (8 582 entries) and maize (1751 entries), is there sufficient sequence information currently available to generate a set of SSR markers suitable for characterizing genetic diversity. We estimate that 30–50 markers with diversity index values >0.5 should be adequate to provide unique genotypes for all accessions in moderately large germ plasm collections. Database searching alone is unlikely to yield sufficient SSR markers in any plant species for mapping or breeding applications (>500 markers). However, cDNA and other sequencing projects are rapidly contributing additional sequence information to public databases for certain species, thus allowing more SSR loci to be found by database searches.

Estimates of the number of SSR markers required for characterizing germ plasm collections or for providing cultivar identification are based on the combined discrimination power provided by multiple markers. This discrimination ability is dependent on the assumption that alleles of one marker are not linked to alleles of other markers across any significant sub-fractions of a collection (linkage disequilibrium). Provided that no linkage disequilibrium occurs, then the discrimination power of the assay increases exponentially with each additional marker. For example, if each marker has a diversity of 0.5 (which corresponds to a 50% chance of discriminating two unrelated plants), then the total discrimination power can be calculated with the formula $X=[1/(1-D)]^n$, where X is the number of unique genotypes that can be generated (equivalent to discrimination power), D is the average diversity of the markers, and n is the number of marker loci assayed. Therefore, a single marker will provide $(1/0.5)^1=2$ genotypes, ten markers give $(2)^{10}=1024$ genotypes, and 20 markers will yield $(2)^{20}>1$ million genotypes, which should provide more than enough discrimination power to categorize all of the accessions in any germplasm collection. Since SSRs frequently have diversity values greater than 0.5, fewer than 20 markers should be required for an informative genotyping of a collection (only 12 markers with a diversity value of 0.7 can generate over 1 million different genotypes).

The value of SSR-specific primer pairs developed in one species to assay polymorphism in another species (heterologous primer pairs) is difficult to predict, but our results are discouraging. The screening of 67 maize and ten seashore paspalum primer pairs on sorghum lines indicated that only 2–3% were suitable for use as markers in sorghum. Also, each primer required individual attention in order to optimize amplification conditions and to verify that polymorphic products contained SSR sequences.

Different results may be found with other combinations of species. The usefulness of primers across species may be a function of relationship between the species; however, the establishment of broadly applicable standards for measuring such relationships in different taxa is difficult. Several investigators have shown that SSR-specific primer pairs developed for soybean, rice, grape (*Vitis vinifera* L.), *Brassica* spp., and barley can be used among closely related species (Akkaya et al. 1992; Lagercrantz et al. 1993; Thomas and Scott 1993; Zhao and Kochert 1993; Saghai Maroof et al. 1994; Kresovich et al. 1995), but the application of these markers across plant genera remains a largely unexplored area. Zhao and Kochert (1993) reported amplification of an SSR-containing product in both maize and bamboo (*Bamboosa vulgaris* Schreber), using a rice PCR primer pair. However, in this experiment, the rice primer pair was developed from an SSR present inside a family of interspersed, highly repeated elements, which were found to be widely conserved among monocot species. Several experiments have shown that primers flanking SSR loci can be used across related genera of mammals, including cetaceans (Schlotterer et al. 1991), rodents (Kondo et al. 1993), and primates (Moore et al. 1991).

At this point, the data indicate that if primers are not available from closely related species, then screening a large number of primers may be necessary in order to find a set that can be used as reliable polymorphic markers in a species of interest. If heterologous primer pairs are readily available from collaborators working on other species, then a large-scale screening may be cost effective. If primers must be designed and synthesized based on sequence information from other species, then this will probably not be an efficient approach.

Despite the effort required, library screening was used in this study as the primary method for developing the SSR markers required for sorghum germplasm analysis. In order to improve the efficiency of the library screening process, a lambda phage library was cloned in the vector LambdaZap II. The phage library allowed the screening of approximately 2000 clones per 150-mm Petri dish, by direct plaque hybridization. Hybridizations were conducted with a mixture of radiolabeled oligonucleotides complementary to the most common di-, tri-, and tetra-meric repeat motifs, in order to identify the maximum number of SSR-containing clones in a single screening.

Library screening is inherently an inefficient process. In the initial screening, approximately 0.2% of library clones hybridized to the mixed SSR probe. A second round of screening was performed on plasmids excised from positively hybridizing plaques, thus greatly reducing the effort wasted by sequencing false positives. Approximately 70% of the sequenced clones were not used for the design and synthesis of primers due to insufficient flanking sequence, clones that contained too few tandem repeats (<14 bp), or severely imperfect repeats. Finally, 65% of primer pairs did not produce polymorphic products from an array of 17 sorghum lines.

Some primer pairs that were monomorphic in the array of 17 cultivars did detect polymorphisms in the set of four genetically dissimilar sorghum lines. In addition, the nondenaturing polyacrylamide gels that were used to identify PCR product size polymorphisms are not capable of resolving differences of one or two base pairs. Work is in progress in our laboratory to characterize the variation in SSR-specific PCR product sizes from a range of sorghum germplasm, using the much more precise technology of semi-automated fluorescent detection in sequencing gels. Primers that were scored as monomorphic in the test array may provide useful polymorphisms in other sorghum germplasm.

Diversity index values were calculated for each polymorphic SSR (Table 3). These values can be used as a relative measure of the utility of each SSR as a tool in categorizing sorghum germplasm and for a general evaluation of the informativeness of SSRs in comparison with other types of molecular markers. However, these diversity values are a direct result of our choice of lines used in the test array, and will not accurately reflect the levels of marker variability across all sorghum germplasm.

The ultimate goal for this sorghum SSR marker-development project is to produce sets of primers that can be used together to assay 4–8 polymorphic loci with a single reaction and a single lane of an electrophoresis gel. This "single-tube multiplex" PCR approach would greatly reduce the cost, improve the speed, and allow the large-scale application of molecular-marker analysis. However, the development of such a high-throughput multiplex system requires rigorously standardized assay conditions. All primer pairs must anneal at the same temperature. None of the primer sequences in a reaction should have substantial homology with each other or with any of the fragments produced by the amplifications. Primer pairs producing inconsistent or difficult-to-score products and loci that have low levels of polymorphism should not be used. Ideally, all of the SSR markers used to characterize plant genotypes should be unlinked. With these constraints, a much larger number of SSRs must be identified, sequenced, and evaluated than the final number required to characterize sorghum germplasm.

Acknowledgements The authors thank C. A. Jester for exceptional work in DNA sequencing and oligonucleotide synthesis, W. B. Chen for DNA extraction and initial library construction, A. L. Westman for many helpful comments on the manuscript, K. B. Foglia for editorial assistance, and E. C. L. Chin at Pioneer Hi-Bred for primer synthesis and other collaboration. We also thank the NC Biotechnology Center for support to M. L. Senior through Institutional Development Grant No. 9110-IDG-1022 and the USDA for NRICGP Grant No. 94-37300-0296 to M. L. Senior. F. G. was supported by project PB93-0350 from DGICYT (Ministerio de Educacion y Ciencia, Spain) and by a fellowship from M. E. C., Programa de Formacion del Profesorado.

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132: 131–139

Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB (1995) Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci 35:1439–1445

Altschul, SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J Mol Biol 215:403–10

Ammer H, Schwaiger F-W, Kammerbauer C, Arriens A, Lazary S, Epplen JT (1992) Exonic polymorphism versus intronic hypervariability in DRB genes: evolutionary persistence and group-specific organization in simple repeat sequences. Immunogenetics 35:330–337

Bell CS, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of arabidopsis. Genomics 19:137–144

Condit RC, Hubbell SP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34:66–71

Feinberg AP, Vogelstein B (1983) A technique for labeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

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

Karper RE (1949) Registration of sorghum varieties V. Agron J 41:536–540

Karper RE (1953) Registration of sorghum varieties VI. Agron J 45:322–323

Kondo Y, Mori M, Kuramoto T, Yamada J, Beckmann JS, Simon-Chazottes D, Montagutelli X, Guenet JL, Serikawa T (1993) DNA segments mapped by reciprocal use of microsatellite primers between mouse and rat. Mammal Genome 4:571–576

Kresovich S, Szewc-McFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterization of simple sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (Rapeseed). Theor Appl Genet 91:206–211

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–1115

Lavi U, Akkaya M, Bhagwat A, Lahav E, Cregan PB (1994) Methodology of generation and characteristics of simple sequence repeat DNA markers in avocado (*Persea americana* M.). Euphytica 80:171–177

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Liu ZW, Jarret RL, Kresovich S, Duncan RR (1995) Characterization and analysis of simple sequence repeat (SSR) loci in seashore paspalum (*Paspalum vaginatum* Swartz). Theor Appl Genet 91:47–52

Mellersh C, Sampson J (1993) Simplifying detection of microsatellite length polymorphisms. BioTechniques 15:582–584

Moore SS, Sargeant LL, King TJ, Mattick JS, Georges M, Hetzel DJS (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10:654–660

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

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Saghai-Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91:5466–5470

Schlötterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. Nature 354:63–65

Scheuring JF, Miller FR (1978) Fertility restorers and sterility maintainers to the milo-kafir genetic cytoplasmic male-sterility system in the sorghum world collection. Texas Agric Expt Stn MP-1367

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36:884–889

Senior ML, Chin ECL, Smith,JSC (1995) Simple sequence repeats in maize – a progress report. Maize Genet Coop Newslett 69:119–120

Stephens JC, Miller FR, Rosenow DT (1967) Conversion of alien sorghums to early combine genotypes. Crop Sci 7:396

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites. Theor Appl Genet 86:985–990

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1–6

Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. Genomics 7:524–530

Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132:1119–1130

Zhao X, Kochert G (1993) Phylogenetic distribution and genetic mapping of a $(GGC)_n$ microsatellite from rice (*Oryza sativa* L.). Plant Mol Biol 21:607–614