Development of Simple Sequence Repeat Markers from Bacterial Artificial Chromosomes without Subcloning

BioTechniques 31:355-362 (August 2001)

ABSTRACT

Simple sequence repeats (SSRs) were isolated from pearl millet bacterial artificial clones (BACs) without any subcloning steps. SSR sequences were targeted using 3' end-anchored SSR primers. Flanking sequences were isolated by suppression PCR. In this pilot study, 25 SSR markers have been developed from 40 BAC pools, comprising a total of 384 clones. This novel way to develop new markers has the added advantage that mapping the SSR markers will anchor individual BACs to the genetic maps and, thus, facilitate the construction of BAC contigs.

INTRODUCTION

Simple sequence repeats (SSRs) or microsatellite DNA sequences are abundant in eukaryotic genomes (9). The number of repeat units is highly variable, and the flanking sequences are relatively conserved among haplotypes of the same species. Length polymorphism relies on a difference in the number of repeat units in different genotypes and is revealed after PCR amplification with primer pairs designed to the SSR flanking regions. This PCR-based SSR marker is generally co-dominant and highly polymorphic, providing an efficient marker sys-

tem for genomic analysis in many species (2,4-7). SSR sequences have generally been identified either in DNA sequence databases or following the screening and sequencing of small-insert genomic libraries. Recently, Cregan et al. (3) reported the development of SSR markers by subcloning a limited number of bacterial artificial clones (BACs). In this paper, we describe a PCR-based method for the development of SSR markers from BACs without subcloning.

This method relies on the suppression PCR technique developed by Siebert et al. (8). Key to the technology is a specifically designed adapter consisting of uneven strands. The shorter bottom strand carries an amine group at its 3' end, which blocks polymerase catalyzed extension of this adapter strand. When a DNA template is ligated to adapters, binding sites for adapter primers are only generated following extension from a template-specific primer. In addition, any generated PCR products that contain double-strand adapter sequences at both ends will form "pan-handle" structures upon denaturation because of the presence of inverted terminal repeats. These structures are more stable than the primer-template hybrids and will not be amplified.

MATERIALS AND METHODS

Restriction-Ligation of BAC DNA

Pearl millet BAC clones stored in a 384-well microplate were pooled in two dimensions. DNA was extracted from the 40 BAC pools by alkaline lysis. BAC DNA (50 ng) were restricted with a frequent cutter, *TaqI*. The majority of DNA fragments were shorter than

1 kb. Ligation to a suppression PCR adapter was carried out in a total volume of 25 μ L One-Phor-All PLUS buffer (Amersham Pharmacia Biotech, Piscataway, NJ, USA) containing 25 μ M adapter, 1 mM ATP, and 0.5 U T4 DNA ligase at 37°C for 3 h. The restriction/ligation mixtures were diluted 10 times with TE buffer and stored at 4°C or -20°C. The adapter was derived from Siebert et al. (8), with minor modifications to fit to the cohesive ends of *TaqI*-restricted DNA fragments (Figure 1).

Targeting SSR Sequences by Suppression PCR

To specifically amplify SSR flanking sequences, we designed SSR primers that were anchored at the 3' end with one selective nucleotide (10). In this pilot study, only two, (AC)₁₀C and (AC)₁₀T, of the six possible anchored CA primers, (AC)₁₀C, (AC)₁₀G, (AC)₁₀T, (CA)₁₀A, (CA)₁₀G, and (CA)₁₀T, were used to target the flanking sequences of CA/GT repeats in the 40 BAC DNA pools. Suppression PCR was performed in 20 μ L HotStar-Taq™ DNA polymerase buffer (Qiagen, Valencia, CA, USA) containing 5 μ L diluted restriction/ligation mixture, 0.5 U HotStarTaq DNA polymerase (5 U/ μ L), 30 ng adapter-primer AP1, 30 ng anchored SSR primer, and 200 μ M dNTPs. The PCR program comprised a pre-denaturation step at 95°C for 15 min and 35 cycles of 30 s denaturation at 94°C, 30 s annealing at 61°C, and 1 min extension at 72°C. The PCR products were mixed with 15 μ L formamide loading dye and denatured for 5 min at 95°C before we loaded 5- μ L samples onto a 4% denaturing polyacrylamide gel. The amplified DNA fragments were revealed with silver staining.

Figure 1. A *TaqI*-restricted and ligated BAC-DNA fragment containing CA/GT repeats. The top strand of the adapter is 5'-GTAATACGACTCATTATAGGGC-3', and the bottom strand is 3'-NH₂-CGTNNNNNNNNNTCAGACACACACACACACACANNNNNNNNAGC-5'. AP1 (the first adapter primer) is 5'-GTAATACGACTCATTATAGGGC-3', and AP2 (the second adapter primer) is 5'-ACTATAGGGCAGCGTGGT-3'.

Short Technical Reports

Semi-Nested PCR and Sequencing of PCR Products

Fragments that were common to pools from both dimensions were excised from the gel and crushed in 100 μ L TE buffer. Five microliters of supernatant were used as template for semi-nested PCR using 75 ng adapter primer AP2, 75 ng anchored SSR primer, 1 U *Taq* DNA polymerase (Roche Molecular Biochemicals, Mannheim, Germany), and 200 μ M dNTPs in a total volume of 50 μ L. The PCR conditions were the same as for the suppression PCR, with the exceptions that pre-denaturation was reduced to 3 min and annealing was carried out at 58°C. The PCR products were purified using the Wizard®PCR Preps columns (Promega, Madison, WI, USA) and sequenced using the ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City,

CA, USA), according to the manufacturer's recommendations. AP2 was used as primer in the cycle sequencing reaction.

Isolation and Sequencing of SSR Motifs and Their Second Flanking Sequences

Forward primers (FPs), designed against the obtained sequences, were used in combination with AP1 to amplify the SSR repeats and their second flanking sequences by suppression PCR from the corresponding BAC pools. The PCR products were diluted 250 times, and 5 μ L were used for semi-nested PCR with the primers FP and AP2. The protocols for suppression PCR, semi-nested PCR, PCR product purification, and DNA sequencing were the same as previously described. Cycle sequencing was carried out with either AP2 or FP.

Confirmation and Testing of the Developed SSR Markers

Reverse primers (RPs) were designed against the second flanking sequences and used in combination with the corresponding FPs to amplify the SSR sequences. In an initial screen, BAC clones from which the SSR originated, restricted and adapter-ligated BAC pool DNA, and genomic DNA from Tift 23DB, which is the inbred line used for the construction of the BAC library, were used as template. SSR markers were then tested for their ability to reveal polymorphism in a panel of 20 pearl millet lines or cultivars. PCR amplification was conducted in a 20- μ L reaction volume with 1 $\times$ reaction buffer containing 50 ng genomic DNA, 5 μ L denatured BAC clone cultures, or 5 μ L restriction/ligation mixtures, 30 ng FP, 30 ng RP, 200 μ M dNTPs, and 0.5 U *Taq* DNA polymerase. The PCR was as fol-

lows: pre-denaturation at 94°C for 3 min, and 35 cycles each of denaturation at 94°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 1 min. PCR products were separated on 5% denaturing polyacrylamide gels and revealed with silver staining.

RESULTS AND DISCUSSION

Suppression and Semi-Nested PCR

Figure 2 shows the results of suppression PCR on the pooled DNA from BACs in lanes A–P and 1–24 from a 384-well microplate. A highly specific amplification was achieved with the anchored SSR primers and the suppres-

sion PCR adapter primer (Figure 1). The presence of common fragments in the X and Y pools indicates the BAC clone from which the fragment was amplified. In some cases, more than two bands were identified (Figure 2, bands 5 and 9). This may be due to the presence of overlapping BAC clones in the pools or to co-migration of different DNA fragments. In this study, only common fragments that were present once in each of the pools and whose origin could be determined unambiguously were excised for further sequencing.

Excision and re-amplification by semi-nested PCR of 27 (AC)₁₀C and 24 (AC)₁₀T fragments, which were common to X and Y BAC pools, produced 23 and 22 clear single bands, respectively (data not shown). Nonspecific

fragments generated during the suppression PCR could be observed in some of the BAC pools, but these failed to amplify during the subsequent semi-nested PCR. Fragments with sizes between 200 and 700 bp were excised and used for sequencing. In total, 40 readable sequences were obtained. Two sequences derived with the (AC)₁₀C anchor primer were identical to sequences generated from the (AC)₁₀T anchor primer. Therefore, 38 FPs were designed. The ideal region for FP design was 200–100 bases from the SSR motif.

Isolation and Sequencing of SSR Motifs and Their Second Flanking Sequences

SSR motifs and their second flanking sequences were targeted using the FP. Most FPs amplified clear single bands. Occasionally, two or three bands were generated, and all clear bands were excised from low-melting agarose. Fragments shorter than 200 bp were discarded because they were expected to yield insufficient flanking sequence for the RP design. In total, 25 RPs were designed against the 25 high-quality sequences containing at least 5 CA/GT repeat units (Table 1).

SSR Markers

Table 1 presents the primer sequences, SSR repeat length and expected product size in Tift 23DB, which is the inbred line used for the construction of the BAC library. All 25 primer pairs efficiently amplified fragments from the microsatellite-originated BAC clones, the restricted and adapter-ligated BAC pool DNA, genomic DNA from Tift 23DB, and three further pearl millet lines (Figure 3). The polymorphism levels of the developed SSR markers were evaluated in 20 selected pearl millet cultivars or lines (Figure 4), and the polymorphism information content (PIC) values varied from 0 to 0.83. The mapping of the developed SSR markers will anchor these BAC clones to the pearl millet genetic map (data not shown).

About 46% (11/24) of the sequences contained less than eight SSR repeat units. They generally produced monomorphic PCR products or displayed very low PIC values. Longer micro-

Figure 2. The suppression PCR products obtained following amplification with AP1 and (AC)₁₀C separated on a 4% denaturing polyacrylamide gel and revealed with silver staining (partial image only). Lanes A–P are the 16 BAC pool DNA samples bulked by rows, and lanes 1–24 are the 24 BAC pool DNA samples bulked by columns. A 1-kb DNA size marker (Life Technologies, Rockville, MD, USA) was added in each lane to form an internal DNA size reference for easy identification of DNA fragments in the two pools. Fragments common to both pools are numbered identically.

Figure 3. Confirmation of the SSRs. Amplification of SSRs, PSMP2201 (A) and PSMP2208 (B), from lane 1, individual BAC clones (bacterial culture); lane 2, restriction/ligation mixture of the corresponding 2-D BAC pool; lane 3, Tift 23DB; lane 4, 81B; lane 5, ICMP 451; and lane 6, 841B. Lane M shows 1-kb DNA size markers (Life Technologies). The primer pair for PSMP2208 amplified double bands and ICMP451 is heterozygous at this locus.

Table 1. SSR Marker Names, Primer Pairs, Repeat Length, and PIC Values

Marker	Derived BACs	Primer Pair (FP/RP)	PCR Product	Repeat Type and Length	PIC Value ^a (No. of Alleles)
PSMP2201	B320O08	5'-CCCCGACGTTATGCGTTAAGTT/TCCATCCATCCATTAATCCACA-3'	364	(GT) ₆	0.69 [4]
PSMP2202	B320D05	5'-CTGCCCTGTTGAGAATAAATGAG/GTTCCGAATATAGAGCCCAAG-3'	161	(GT) ₈	0.42 [2]
PSMP2203	B320O12	5'-GAACTTGATGAGTGCCACTAGC/TTGTGTAGGGAGCAACCTTGAT-3'	357	(GT) _{18 imp.b}	0.8 [6]
PSMP2204	B320A10	5'-TGCCCTTCCTGACTATGTTTCC/AGATATGGCGAACGTGAGGAG-3'	266	(GT) ₇	0.1 [2]
PSMP2205	B320B17	5'-AGGTGCTCACGAGCTGTAAGAG/AGCAAGACACTATTTTACCATC-3'	202	(GT) _{7c}	0.0 [1]
PSMP2206	B320D02	5'-AGAAGAAAGAGGGGGTAAGAAAGGAG/AGCAACATCCGTAGAGGTAGAAG-3'	203	(GT) ₁₁	0.83 [6]
PSMP2207	B320D05	5'-CAGGGCATACTTCAAGATTGATTC/GTCCACTTGTATTCTCTATCAC-3'	304	(GT) ₅	0.0 [1]
PSMP2208	B320E09	5'-GAAAGAGCAAACTGAACAATCCC/ACTTTGCCCTGGATGATCCTC-3'	253	(GT) ₁₀	0.78 [5]
PSMP2209	B320F04	5'-TTGGACGATTTGGAAGCATAG/GAGGAAAAGAGCCATACAGAGAC-3'	334	(GT) ₆ , (CT) ₇	0.62 [3]
PSMP2210	B320F05	5'-CAATGATGACCGTAATCTGGGTG/GGGCAAGATATGTGAATCAAG-3'	313	(GT) ₁₂	multiple loci
PSMP2211	B320I09	5'-CTGCATGACGTGTACCAATACC/AACAAATCAGCACAGCCTCC-3'	244	(GT) ₆	0.38 [2]
PSMP2212	B320N06	5'-GATTGGATGGCAGTGTCTGG/CAAAACCAGCCATCAACAACCAAG-3'	191	(GT) ₆	0.0 [1]
PSMP2213	B320N06	5'-CCCCAAAAGAACCCACACCCAC/GTTGATGCTACTGCTGTTTG-3'	197	(GT) ₁₀	0.1 [2]
PSMP2214	B320O18	5'-CGCACAGTACGTGTGAGTGAAG/GATTGAGCAGCAAAAACCCAGC-3'	246	(GT) ₉	0.74 [4]
PSMP2215	B320P12	5'-CCACGTCATTAGAGTAATCCGAG/ACTCAAATCCCAATCTTGAATC-3'	238	(GT) ₆	0.48 [4]
PSMP2216	B320D17	5'-GTATGTAGGTGATGGTTAGTTCG/AGTAGAAGGATGGAGTCATACAG-3'	140	(GT) _{7 imp}	0.0 [1]
PSMP2217	B320F12	5'-TTGTCCTCTCCGGCCAGATCCT/GCGTGTATGTGTTGGTACGCGAT-3'	141	(GT) ₈	0.0 [1]
PSMP2218	B320H19	5'-CTCTGTAAGTTCCTGGTGCTCAA/TCAGGCCAGTAACACATCTCAA-3'	250	(GT) _{12(GA)₃₀}	0.82 [9]
PSMP2219	B320I12	5'-ACTGATGGAATCTGCTGTGGAA/GCCCCGAAGAAAAGAGAACATAGAA-3'	254	(GT) ₇	0.34 [3]
PSMP2220	B320I12	5'-GCATCCTTCACCATTCAGACA/TGGGAAACAGATGGAGAAAAGAG-3'	128	(GT) ₁₁	0.66 [4]
PSMP2221	B320J21	5'-TTGCCGTCAGCAATGTGCCT/CCGAAAGTGCCCAAGTGCCCAA-3'	203	(GT) ₅	multiple bands
PSMP2222	B320J21	5'-TGGCTTCCAGACTAATCATCAC/TTATTTAGCGCGGAGATTGAC-3'	155	(GT) ₆	0.1 [2]
PSMP2223	B320L06	5'-CATGCTTCTTCTCTTTTGTAACC/CAGCTCTTTGATCTCACTACAC-3'	190	(GT) ₇	0.0 [1]
PSMP2224	B320L19	5'-GGCGAAATTGGAATTCAGATTG/CGTAATCGTAGCGTCTCGTCTAA-3'	155	(TG) ₁₀	0.58 [3]
PSMP2225	B320N23	5'-CCGTACTGATGATACTGATGGTT/TGGGAGGTAAGCTCAGTAGTGT-3'	238	(GT) ₁₂	multiple loci

^aThe polymorphism information content (PIC) was estimated according to Anderson et al. (1):

$$PIC_i = 1 - \sum_{j=1}^n p_{ij}^2$$

where p_{ij} is the frequency of the j th pattern of marker i , and the sum is made over n patterns, assuming homozygosity of the pearl millet inbred lines.

^bImp. indicates imperfect repeat.

^cRepeat length is unknown because the sequencing reaction was unable to proceed through the repeats.

Short Technical Reports

Figure 4. SSR profile generated with PSMP2203. Lanes 1–20 are the selected 20 pearl millet cultivars or lines. Lane M shows 1-kb DNA size markers (Life Technologies).

satellite sequences (i.e., more than eight repeats) generally displayed a high level of polymorphism among the 20 pearl millet varieties tested. In this pilot study, suppression PCR was

carried out with HotStarTaq DNA polymerase in 1× HotStarTaq DNA polymerase buffer at an annealing temperature of 61°C. It is expected that the use of a more stringent annealing temperature, from 63°C to 65°C, may lead to the isolation of longer microsatellites.

Comparison of the isolated microsatellite sequences with those of the 3'-anchored SSR primers showed some mismatch with the 3' anchor (data not shown). Therefore, it is possible that the same microsatellite is amplified by differentially anchored SSR primers. In this study, two duplicate sequences were identified with (AC)₁₀C and (AC)₁₀T. To avoid the isolation of duplicated SSR sequences, the use of two or three anchor nucleotides may be required at the 3' end. Alternatively, a mixture of the six 3'-end SSR anchor primers can be used in the suppression PCR.

REFERENCES

1. Anderson, J.A., G.A. Churchill, J.E. Autrique, S.D. Tanksley, and M.E. Sorrells. 1993. Optimizing parental selection for genetic linkage maps. *Genome* 36:181-186.
2. Bryan, G.J., A.J. Collins, P. Stephenson, A. Orry, J.B. Smith, and M.D. Gale. 1997. Isolation and characterization of microsatellites from hexaploid bread wheat. *Theor. Appl. Genet.* 94:557-563.
3. Cregan, P.B., J. Mudge, E.W. Fickus, L.F. Marek, D. Danesh, R. Denny, R.C. Shoemaker, B.F. Matthews et al. 1999. Target isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. *Theor. Appl. Genet.* 98:919-928.
4. Dib, C., S. Fauré, C. Fizames, D. Samson, N. Drouot, A. Vignal, P. Millasseau, S. Marc et al. 1996. A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature* 380:152-154.
5. Dietrich, W.F., J. Miller, R. Steen, M.A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, M.J. Daly et al. 1996. A comprehensive genetic map of the mouse genome. *Nature* 380:149-152.
6. Hudson, T.J., L.D. Stein, S.S. Gerety, J.L. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista et al. 1995. An STS-based map of the human genome. *Science* 270:1945-1954.
7. Senior, M.L., J.P. Murphy, M.M. Goodman, and C.W. Stuber. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Sci.* 38:1088-1098.
8. Siebert, P.D., A. Chenchik, D.E. Kellogg, K.A. Lukyanov, and S.A. Lukyanov. 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* 23:1087-1088.
9. Tautz, D. and M. Renz. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* 12:4127-4138.
10. Zietkiewicz, E., A. Rafalski, and D. Labuda. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20:176-183.

This manuscript is an output from a project funded by the UK Department for International Development (DFID) for the benefit of developing countries. The views expressed are not necessarily of those of DFID. John Innes Centre is subsidized by the Biotechnology and Biological Sciences Research Council, UK. Address correspondence to Dr. Xiaoquan Qi, John Innes Centre, Colney, Norwich NR4 7UH, UK. e-mail: xiaoquan.qi@bbsrc.ac.uk

Received 13 September 2000; accepted 24 January 2001.

X. Qi, S. Lindup, T.S. Pittaway, S. Allouis, M.D. Gale, and K.M. Devos
*John Innes Centre
 Norwich, UK*