
Plant Molecular Biology Reporter **16:** 341–349, 1998.

Protocols

A High Through-put Procedure for Capturing Microsatellites from Complex Plant Genomes

JAMES P. CONNELL, SUJATA PAMMI[1], MUHAMMAD J. IQBAL, TIM HUIZINGA and AVUTU S. REDDY*

Crop Biotechnology Center and Dept. of Soil and Crop Sciences, [1]Crop Biotechnology Center and Dept. of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843-2123, USA

Abstract. A method is outlined for large-scale isolation and characterization of microsatellite sequences from complex plant genomes. The method presented here differs from the previously published procedures in the use of randomly sheared (nebulized) genomic DNA for adapter-ligation, rigorous removal of biotinylated oligos, and high-density colony blots for constructing enriched libraries. Using this method we have constructed cotton microsatellite enriched libraries with over 20% (high stringency screening) or 75% (by random sequencing). Thus far we have identified and sequenced over 500 cotton microsatellites using this procedure. The procedure can be used to generate enriched SSR libraries from genomic DNA in about one week. High throughput screening and automated DNA sequencing can be accomplished in less than one month.

Key words: cotton, DNA markers, nebulization, simple sequence repeats

Introduction

Microsatellites, also referred to as Simple Sequence Repeats (SSRs), are stretches of DNA containing tandemly repeating di-, tri-, or tetra nucleotide units ubiquitously distributed throughout the eukaryotic genomes. SSRs are becoming the standard DNA markers for plant and animal genome analysis. A wide variety of methods for construction of libraries enriched for microsatellite sequences have been reported, including those based on vectorette PCR using anchored primers (Lench et al., 1996), hybridization of adapter-ligated DNA to primer-bound filters (Edwards et al., 1996), extension of

*Author for correspondence. Present address: Laboratory for Crop Genome Analysis, Crop Biotechnology Center, Texas A&M University, College Station, Texas 77843-2123. e-mail: sam@bio.tamu.edu; fax: (409) 862-4790.

uracil-containing DNA templates using repeat primers (Ostrander et al., 1992) and hybridization of PCR generated single stranded template to biotinylated oligos (Kijas et al., 1994). We have chosen a method based on that of Prochazka (1996) that employs size-fractionated adapter-ligated template with biotinylated oligonucleotide capture. The modified method relies on nebulization of genomic DNA and a low number of thermocycling steps to obtain usable 5′ and 3′ flanking sequences of microsatellite loci and high-density colony blot screening.

Upland cotton, *Gossypium hirsutum* L., is an allotetraploid ($2n = 4x = 52$) with a genome size of 2300 Mb and 4.5 pg/2x DNA content (Arumuganathan and Earle, 1991). The level of Restriction Fragment Length Polymorphism (RFLP) within cotton races is low (Reinisch et al., 1994; Dr Russell Kohel, USDA/ARS, personal communication), which makes the study of genetic diversity and pedigree analysis difficult. Therefore, a new type of co-dominant hypervariable marker system such as SSRs is required for successful application of DNA markers in cotton. The method described here has been used to isolate cotton microsatellites but it is readily applicable to other plant species.

Materials and Methods

G. hirsutum L., var. Coker 312 seed was obtained from Dr John Gannaway, Texas Agricultural Experiment Station, Lubbock and grown in the greenhouse of the Crop Biotechnology Center. Young leaves were harvested and genomic DNA was extracted according to Paterson et al. (1993) and quantified using a fluorimeter (Hoefer-Scientific).

Nebulization

- Combine 10 μg genomic DNA, 0.2 ml 10× TM buffer (0.5 M Tris-HCl [pH 8.0], 0.15 M $MgCl_2$), 1.25 ml 80% glycerol and 0.535 ml H_2O in the nebulizer[1].
- Nebulize for 2.5 min in an ice-cold water bath at 10–30 psi[2].
- Centrifuge entire unit in a tabletop centrifuge at 2500 rpm. Distribute sample into four 1.5 ml microfuge tubes, add 0.5 volume of 7.5 M NH_4Oac and 2 volumes of ethanol. Precipitate 30 min at −20 °C.
- Centrifuge, dry the pellets and resuspend in 40 μl of 1× TM buffer and pool DNA from all four tubes. Examine 2 μl on agarose gel to confirm that desired size has been achieved before proceeding with end repair.

- Polish ends by mixing 20 μl (approximately 5 μg) DNA fragments, 1 μl 2 mM dNTPs, 1 μl T4 DNA polymerase (5 units), 2 μl Klenow fragment (10 units). Incubate 30 min at room temperature. Terminate reaction by heating at 70 °C for 15 min and cool on ice.

Adapter ligation and pre-selective amplification

- Mix 5 μl of nebulized genomic DNA fragments (approximately 1 μg) with 1 μg adapter[3], 4 μl 5× ligation buffer (Gibco-BRL), H_2O to 19 μl and 1 μl of T4 DNA ligase in a 0.5 ml microfuge tube. Ligate overnight at 16 °C, incubate the mixture at 70 °C for 15 min and diluted to 50 μl with H_2O.

- To select for adapter ligated fragments, one microlitre of the diluted ligation mixture is expanded by PCR amplification with AP-11 primer[3] for 15–18 cycles. Combine in a 0.2 ml PCR tube: 36 μl H_20, 5 μl 10× PCR Buffer (Gibco/BRL), 5 μl 2 mM dNTPs, 2.5 μl 50 mM $MgCl_2$, 0.5 μl AP-11 primer (1 μg/μl), 0.25 μl (1.25 units) *Taq* DNA polymerase (Gibco/BRL) and 1 μl of adapter-ligated DNA.

- Amplify using an initial denaturation at 96 °C for 1 min, then 15–18 cycles of denaturation at 95 °C for 10 s, annealing at 55 °C for 10 s, and extension at 72 °C for 2.5 min followed by a final soak of 5 min at 72 °C. Check 5–10 μl on an agarose gel. Aliquots of amplified adapter-ligated DNA may be stored at −20 °C.

SSR capture (a modified procedure based on Prochazka, 1996)

- In a thin walled PCR tube, mix 100 ng of adapter-ligated expanded DNA fragments with 0.5 μg each of biotinylated repeat oligo[4] in a final volume of 100 μl 6× SSC/0.1% SDS. Denature samples at 95 °C for 5 min followed by ramping to 60 °C at −2 °C/min in a PE9600 PCR machine (Perkin-Elmer). Allow hybridization to occur at 60 °C for 1 h in the thermocycler.

- Following hybridization, add 200 μg of streptavidin-coated paramagnetic beads (Promega #Z5481) which are equilibrated in 6× SSC and mix by pipetting. Incubate the mixture for 15 min at 60 °C.

- Remove supernatant using a magnetic stand (Promega) and wash the beads 2× at room temperature with 6× SSC/0.1% SDS, 2× at 60 °C with 6× SSC/0.1% SDS and 2× at room temperature with 6× SSC. Washes are for 15 min each.

- Elute captured DNA by addition of 50 μl 0.1 N NaOH at 60 °C for 5 min, followed by 50 μl of 1 M Tris-HCl (pH 8.0).
- Desalt by centrifugation through a Sephadex G-50 spun column.
- Remove[5] residual biotinylated oligos by either fractionation through low melting- point agarose or multiple (3 or more) low speed spins through Sephadex G-50 spun columns with spins of 2 min or less.
- Amplify one microliter of the selected fragments using AP-11 primer as described above.

Library construction and archiving

- Ligate one microliter of amplified, selected material with 2 μl pCRII vector (Invitrogen) in a final volume of 10 μl according to manufacturer's recommendations. Ligate overnight at 16 °C.
- Inactivate by heating at 65 °C for 15 min. Transform *E. coli* DH10B cells by electroporation. Plate on YT-agar containing 100 μg ampicillin/ml. Plates are treated with X-GAL and IPTG for blue/white colony selection (Sambrook et al., 1989). Grow overnight at 37 °C.
- Pick white colonies and archive into 384-well microtitre plates (Nalge Nunc International #242765) containing 45 μl per well of plate medium (45 ml LB, 5 ml 10× freezer medium, 50 μl ampicillin stock [100 mg/ml]). Freezer medium is 0.36 M K_2HPO_4, 0.132 M KH_2PO_4, 0.017 M NaCitrate, 0.004 M $MgSO_4$, 0.068 M $(NH_4)_2SO_4$, and 44% glycerol.

High density colony filter preparation and screening

- Four 384-well microtitre plates may be used to prepare one 8 × 12 cm size HybondN$^+$ (Amersham #RPN303B) filter in a 3 × 3 grid by duplicating each plate and leaving a 'blank' in the center. Spot using a Beckman Biomek 2000 Automated Laboratory Workstation interfaced with a 384-pin tool[6].
- Grow spotted colonies overnight at 37 °C on filters placed on YT-agar media containing 100 μg ampicillin/ml.
- Process filters to lyse bacteria and fix nucleic acids to the filter. First incubate filter on a piece of Whatman 3 M paper impregnated with 10% SDS for

4 min, then transfer to a second piece of paper soaked in 0.5 N NaOH/1.5 M NaCl for 7 min, then transfer to another piece of paper containing 0.5 M Tris-HCl (pH 7.5)/1.5 M NaCl for 5 min. Transfer to 2× SSC-containing filter paper for 3 min, followed by air drying[7].

- Screen filters by hybridization at 60 °C overnight using end-labeled (^{32}P) SSR-oligonucleotides as probe (Sambrook et al., 1989). Wash at room temperature using 1× SSC/0.1% SDS and 0.3× SSC/0.1%SDS. Catalogue positive clones and assign number designations according to position on the grid. Perform minipreps using QIAGEN Spin MiniKit (#27106) and quantify by spectrophotometric measurement. To simplify sequencing, sample concentrations are adjusted to 250 ng/μl.

Sequencing and analysis of microsatellite loci

- Set up cycle sequencing reaction[8] which contains 1 μl 35% DMSO, 2 μl BigDye Terminator Mix (PE Applied Biosystems #4303152), 1 μl primer (M13 forward or M13 reverse at 10 pmoles/μl), 250 ng of template DNA and adjust total volume to 7 μl with water.
- Cycle sequence for 30 cycles of 96 °C for 10 s, 50 °C for 5 s, 60 °C for 4 min.
- Add 7 μl of H_20 to each sample using a multichannel pipette.
- Set up spin plates containing 175 μl Sephadex G-50 per well using Millipore MAHA N45 plates affixed to a microtitre collection plate (Corning Thermowell #6511). Centrifuge at 2500 rpm for 4 min.
- Discard the flow-through and affix a new microtitre collection plate. Apply samples using a multichannel pipette. Centrifuge at 2500 rpm for 4 min.
- Dry samples in a speed-vac. Resuspend in 2 μl of loading dye and denature using a 96-well format heat block.
- Load samples on an ABI377 automated DNA sequencer and collect sequence information.

Notes

1. Nebulizers (IPI Medical products Inc., Chicago; product #4207) are modified by removing the cylinder drip ring, trimming off the outer rim, and placing it back in the nebulizer in an inverted position. The large hole in the top cover is sealed with a plastic stopper. The small hole in the top cover is connected to N_2 gas by Tygon tubing.

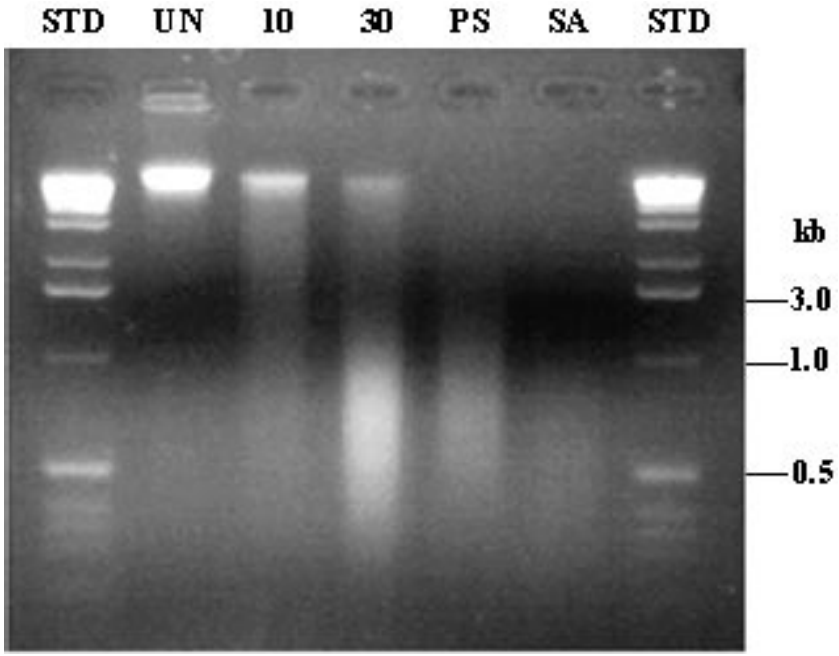

Figure 1. Cotton genomic DNA (lane UN) was nebulized at 10 psi (Lane 10) and 30 psi (lane 30). The 30 psi fraction was adapter-ligated and pre-selective amplification was performed (lane PS). Biotin-captured microsatellite fraction was selected and amplified (Lane SA). Representative molecular size standards (STD) are given in kb.

2. One needs to do a pilot study to find time and pressure requirements to obtain desired size range (see Figure 1).
3. Mix equal amounts of AP-11 [5′-CTCTTGCTTAGATCTGGACTA-3′] and AP-12 [5′-TAGTCCAGATCTAAGCAAGAGCACA-3′] primers and heat to 94 °C and then cool slowly to 30 °C. Store this adapter at −20 °C.
4. SSR-containing sequences are selected by hybridization with 5′-biotinylated oligos of the following composition: $[GA]_{20}$, $[CA]_{20}$, $[AGA]_{15}$, $[ACA]_{15}$, $[CAT]_{15}$ and $[CTA]_{15}$.
5. Removal by multiple low speed centrifugation through Sephadex G-50 is more convenient but is not easily assayable for complete primer removal. Millipore Ultrafree-MC 100,000 filter units may also be used, but we find that Millipore Ultrafree-MC 30,000 filter units retain significant amounts of biotinylated oligos.
6. High-density colony filter can also be prepared manually using a Multi-Blot Replicator with a Colony Copier from V&P Scientific, Inc.
7. To enhance signal/noise ratio we do a second round of lysis and neutralization by immersing the filter in 0.4 N NaOH for 15 min, followed by gentle agitation in 5× SSC for 10 min, rinsing in 1× SSC/0.1% SDS for 15 min and air drying.
8. For convenience of handling samples are run in a 96-well format containing multiples of 12 from 24–96 total samples.

Results and Discussion

Our laboratory is interested in developing microsatellite markers for cotton, rice and pecan. Thus it is essential to optimize a procedure that produces reproducible and usable microsatellite sequences in a cost-effective way. In our experience two of the most critical factors that enhance the overall success of this enrichment technique are size selection prior to adapter ligation and complete removal of biotinylated oligonucleotides prior to final expansion by PCR.

Selection of DNA fragments in the range of 0.4–1.0 kb will ensure that positives can be sequenced in one pass, thus reducing overall cost of analysis. For size selection, genomic DNA can be treated by restriction enzyme digestion or sheared by nebulization. For digestion, choose enzymes which do not generate 3′ or 5′ overhangs. Enzymes with either 6- or 4-base recognition sequences may be employed, singly or in combinations. However, it is almost always necessary to fractionate these fragments through agarose gel and elute the desired size range of 0.4–1.0 kb before proceeding with adapter ligation. Nebulization is preferred since it is expected to generate fragments of the desired size range without bias according to sequence context. As depicted in Figure 1, cotton genomic DNA was nebulized according to the above protocol and verified on an agarose gel. It is evident from this that size fractionation is not required if you optimize conditions for each experimental DNA.

Removal of biotinylated oligonucleotides will ensure that truncated PCR products lacking 5′- or 3′-flanking sequence do not predominate. Prior to final expansion, samples must be desalted and the residual primers removed, as high salt inhibits PCR amplification and residual primers generate truncated PCR products that lack flanking sequence. Methods that have proven successful include desalting over Sephadex G-50 columns followed by either (a) several additional low-speed Sephadex G-50 spins for 1 min followed by PCR amplification with AP-11 primer or (b) fractionation over low-gelling temperature agarose. For the latter method, slices are excised from the gel and melted at 70 °C. One-microlitre aliquots are used for PCR amplification with primer AP-11 for 20 cycles.

By this protocol we have successfully isolated and sequenced over 500 unique microsatellites from cotton. Enriched libraries contain up to 224 SSR-containing clones out of 1536 colonies when probed at high stringency (Figure 2). Of the reported enrichment techniques for genomic DNA, the method presented here is most similar to that of Kijas et al. (1994). However, the use of adapter-ligated DNA with minimal cycling reduces the bias in the library towards very small or highly repetitive inserts, which may occur when preparing single stranded templates by the method of Kijas et al. (1994) due to the large number of cycles required for arithmetic expansion.

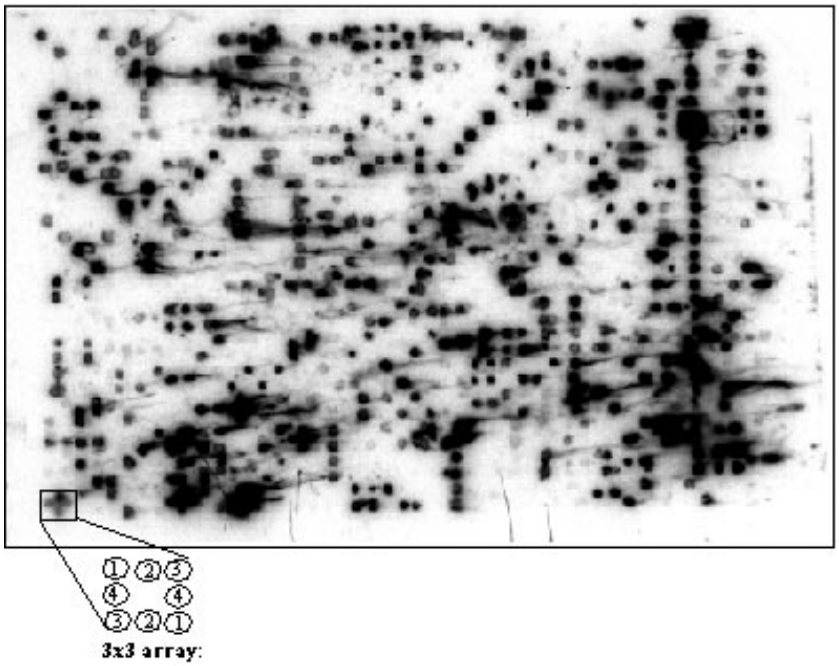

Figure 2. Screening of microsatellite enriched libraries: High density colony blot probed with a $(GA)_{20}$ repeat oligo. Out of 1536 on this filter, approximately 250 hybridized to the repeat probe.

The method presented differs from that of Prochazka (1996) in the use of randomly sheared (nebulized) genomic DNA for adapter-ligation and rigorous removal of biotinylated oligos, which may interfere with the final PCR expansion. Like the methods of Edwards et al. (1996) and Prochazka (1996), libraries may be enriched for several repeat motifs at the same time, provided that the oligos employed do not have widely differing melting temperatures. These authors found 20% positive colonies in the enriched library by high stringency hybridization. Of the other methods, enrichment varied from 40–70% based on high stringency hybridization (Ostrander et al., 1992) or random sequencing (Edwards et al., 1996). It is likely that our estimate of enrichment is lower than the actual value for our libraries, as random sequencing of a CA-enriched library revealed 8 microsats in 12 clones (i.e., 75% enrichment). Sequencing random clones would be expected to yield a higher proportion of short and/or imperfect repeats than those identified by hybridization. None of the clones identified by high stringency hybridization had imperfect repeats according to the definition of Broun and Tanksley (1996) in which a motif contains an uninterrupted stretch of more than 10 dinucleotide repeats.

Library production is rapid, with typical times for genomic DNA isolation to library screening being 2 weeks to 1 month. We have also recently applied this procedure to our pecan and rice projects. Furthermore, we have shared this procedure to other groups working on sorghum, *Arabidopsis* and have proven successful in obtaining enriched SSR libraries.

Acknowledgements

This work was supported by the USDA-NRICGP Plant Genome Grant (#97-35300-4585) to ASR. The authors gratefully acknowledge Dr John Gannaway for providing cotton seed.

References

Arumunganathan K and Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Reptr 9: 208–218.

Broun P and Tanksley SD (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol Gen Genet 250: 39–49.

Edwards KJ, Barker JHA, Daly A, Jones C and Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20: 758–760.

Kijas JMH, Fowler JCS, Garbett CA and Thomas MR (1994) Enrichement of microsatellites from the citrus genome using biotinylated oligonulectide sequences bound to streptavidin-coated magnetic particles. BioTechniques 16: 657–662.

Lench NJ, Norris A, Bailey A, Booth A and Markham AF (1996) Vectorette PCR isolation of microsatellite repeat sequences using anchored dinucleotide primers. Nucl Acids Res 24: 2190–2191.

Ostrander EA, Jong PM, Rine J and Duyk G (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. Proc Natl Acad Sci USA 89: 3419–3423.

Paterson AH, Brubaker C and Wendel JF (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Molecular Biol Reptr 11: 122–127.

Prochazka M (1996) Microsatellite hybrid capture technique for simultaneous isolation of various STR markers. Genome Res 6: 646–649.

Reinisch AJ, Dong J-m, Brubaker CL, Stelly DM, Wendel JF and Paterson AH (1994) A detailed RFLP map of Cotton, *Gossypium hirsutum* X *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138: 829–847.

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, New York.